PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12Q 1/68, C12P 19/34, C07H 21/04

A1

(11) International Publication Number: **WO 98/23779**

(43) International Publication Date: 4 June 1998 (04.06.98)

(21) International Application Number: PCT/US97/21358

(22) International Filing Date: 25 November 1997 (25.11.97)

(30) Priority Data:
60/031,102 26 November 1996 (26.11.96) US

(71) Applicant *(for all designated States except US)*: UNIVERSITY OF UTAH RESEARCH FOUNDATION [US/US]; Suite 170, 421 Wakara Way, Salt Lake City, UT 84108 (US).

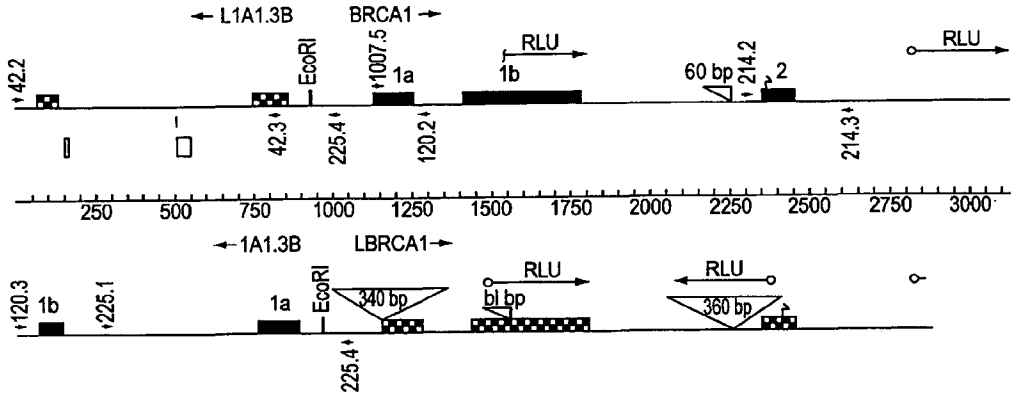
(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: BARKER, David, F. [US/US]; 2768 Comanche Drive, Salt Lake City, UT 84108 (US). LIU, Xudong [CN/US]; Apartment 6, 101 South, 800 East, Salt Lake City, UT 84102 (US).

(74) Agents: IHNEN, Jeffrey, L. et al.; Rothwell, Figg, Ernst & Kurz, Suite 701 East, 555 13th Street N.W., Columbia Square, Washington, DC 20004 (US).

(81) Designated States: CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: THE BRCA1 AND 1A1.3B PROMOTERS ARE PARALLEL ELEMENTS OF A GENOMIC DUPLICATION AT 17Q21



(57) Abstract

The results of experiments aimed at detecting polymorphisms and mutations in the BRCA1 promoter region as well as comparisons of two published DNA sequences indicated that two similar but distinct copies of this region exist in the human genome. PCR primers specific for amplification of each of the two promoter regions were isolated from rearrangement–resistant libraries. Sequence analysis of the clones and specific PCR products reveals two similar genomic rearrangements of head–to–head genes. The BRCA1 gene is closely apposed to a gene structure that is similar but not identical to 1A1.3B and the 1A1.3B gene is apposed to a gene structure that has strong similarity to BRCA1 but also has significant differences. The features of the BRCA1 and 1A1.3B promoter region are shown in the Figure. STS analysis of YAC and P1 clones located in the vicinity of BRCA1 indicates that these similar promoter regions are elements of a direct duplication. New hypotheses for genetic mechanisms that may be involved in breast and ovarian cancer etiology are raised by the identification of this duplicated genetic structure on chromosome 17q. Also presented are polymorphisms in the duplicated genes which polymorphisms are useful in tracking chromosomal rearrangement of these genes.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

TITLE OF THE INVENTION

THE BRCA1 AND 1A1.3B PROMOTERS ARE PARALLEL ELEMENTS OF A GENOMIC DUPLICATION AT 17Q21

This application was made with Government support under Grant Nos. NCI CA63689 and UCI sub-contract J92CA56554 (from NCI CA58660) and NIH CA42014, funded by the National Institutes of Health, Bethesda, Maryland. The United States Government has certain rights in the invention.

FIELD OF THE INVENTION

The present invention relates generally to the field of human genetics. Specifically, the present invention relates to a gene, named LBRCA1 (which stands for "Like BRCA1"), which is very similar to a human breast and ovarian cancer predisposing gene (BRCA1), some mutant alleles of which cause susceptibility to cancer. The invention also relates to a gene called 1A1.3B and a very similar gene named L1A1.3B (for Like 1A1.3B). L1A1.3B is located extremely close to BRCA1 in a head to head configuration while LBRCA1 and 1A1.3B are similarly located very close to each other also in a head to head arrangement, wherein genes that have 5' ends located immediately adjacent to one another are said to be "head-to-head". The BRCA1/L1A1.3B and LBRCA1/1A1.3B regions are a result of gene duplication. Knowledge of the LBRCA1 sequence is important for the analysis of BRCA1 for mutations because the very high similarity between the two genes could lead to problems when trying to analyze BRCA1. Extensive testing of persons for mutations in BRCA1 is expected to begin very soon. The LBRCA1 and L1A1.3B contain promoter regions similar to the promoters for BRCA1 and 1A1.3B. These additional promoters, which are in close proximity to the BRCA1 and 1A1.3B genes, may affect transcription of these latter genes.

A further aspect of the present invention is that the knowledge of the chromosomal arrangement of these genes and the fact that there has been a gene duplication, is useful in looking for mutations, other than mutations directly within BRCA1, which could affect proper transcription of BRCA1 and may be responsible for

breast or ovarian cancer. Another aspect of the invention is that polymorphisms in or near LBRCA1 and L1A1.3B have been found and these are useful in tracking the chromosomal arrangement of these genes as well as BRCA1 and 1A1.3B to determine whether rearrangement has occurred.

The publications and other materials used herein to illuminate the background of the invention, and in particular, cases to provide additional details respecting the practice, are incorporated herein by reference, and for convenience, are referenced by author and date in the following text and respectively grouped in the appended List of References.

BACKGROUND OF THE INVENTION

The genetics of cancer is complicated, involving multiple dominant, positive regulators of the transformed state (oncogenes) as well as multiple recessive, negative regulators (tumor suppressor genes). Over one hundred oncogenes have been characterized. Fewer than a dozen tumor suppressor genes have been identified, but the number is expected to increase beyond fifty (Knudson, 1993).

The involvement of so many genes underscores the complexity of the growth control mechanisms that operate in cells to maintain the integrity of normal tissue. This complexity is manifest in another way. So far, no single gene has been shown to participate in the development of all, or even the majority of human cancers. The most common oncogenic mutations are in the H-ras gene, which is found in 10-15% of all solid tumors (Anderson *et al.*, 1992). The most frequently mutated tumor suppressor genes are the TP53 gene, homozygously deleted in roughly 50% of all tumors, and CDKN2, which was homozygously deleted in 46% of tumor cell lines examined (Kamb *et al.*, 1994). Without a target that is common to all transformed cells, the dream of a "magic bullet" that can destroy or revert cancer cells while leaving normal tissue unharmed is improbable. The hope for a new generation of specifically targeted antitumor drugs may rest on the ability to identify tumor suppressor genes or oncogenes that play general roles in control of cell division.

The tumor suppressor genes which have been cloned and characterized influence susceptibility to: 1) Retinoblastoma (RB1); 2) Wilms' tumor (WT1); 3) Li-Fraumeni (TP53); 4) Familial adenomatous polyposis (APC); 5) Neurofibromatosis type 1 (NF1); 6)

Neurofibromatosis type 2 (NF2); 7) von Hippel-Lindau syndrome (VHL); 8) Multiple endocrine neoplasia type 2A (MEN2A); and 9) Melanoma (CDKN2).

Tumor suppressor loci that have been mapped genetically but not yet isolated include genes for: Multiple endocrine neoplasia type 1 (MEN1); Lynch cancer family syndrome 2 (LCFS2); Neuroblastoma (NB); Basal cell nevus syndrome (BCNS); Beckwith-Wiedemann syndrome (BWS); Renal cell carcinoma (RCC); Tuberous sclerosis 1 (TSC1); and Tuberous sclerosis 2 (TSC2). The tumor suppressor genes that have been characterized to date encode products with similarities to a variety of protein types, including DNA binding proteins (WT1), ancillary transcription regulators (RB1), GTPase activating proteins or GAPs (NF1), cytoskeletal components (NF2), membrane bound receptor kinases (MEN2A), cell cycle regulators (CDKN2) and others with no obvious similarity to known proteins (APC and VHL).

In many cases, the tumor suppressor gene originally identified through genetic studies has been shown to be lost or mutated in some sporadic tumors. This result suggests that regions of chromosomal aberration may signify the position of important tumor suppressor genes involved both in genetic predisposition to cancer and in sporadic cancer.

One of the hallmarks of several tumor suppressor genes characterized to date is that they are deleted at high frequency in certain tumor types. The deletions often involve loss of a single allele, a so-called loss of heterozygosity (LOH), but may also involve homozygous deletion of both alleles. For LOH, the remaining allele is presumed to be nonfunctional, either because of a preexisting inherited mutation, or because of a secondary sporadic mutation.

Two genes that are predisposing for breast cancer have recently been cloned and characterized. These are *BRCA1* (Miki et al., 1994; Futreal et al., 1994) and *BRCA2* (Wooster et al., 1995; Tavtigian et al., 1996). Breast cancer is one of the most significant diseases that affects women. At the current rate, American women have a 1 in 8 risk of developing breast cancer by age 95 (American Cancer Society, 1992). Treatment of breast cancer at later stages is often futile and disfiguring, making early detection a high priority in medical management of the disease. Ovarian cancer, although less frequent than breast cancer is often rapidly fatal and is the fourth most common cause of cancer mortality in American women. Genetic factors contribute to an ill-defined proportion of breast cancer

incidence, estimated to be about 5% of all cases but approximately 25% of cases diagnosed before age 40 (Claus *et al.*, 1991). Breast cancer has been subdivided into two types, early-age onset and late-age onset, based on an inflection in the age-specific incidence curve around age 50. Mutation of one gene, BRCA1, is thought to account for approximately 45% of familial breast cancer, but at least 80% of families with both breast and ovarian cancer (Easton *et al.*, 1993).

There were intense efforts to isolate the BRCA1 gene after it was first mapped in 1990 (Hall *et al.*, 1990; Narod *et al.*, 1991). A second locus, BRCA2, was mapped to chromosome 13q (Wooster *et al.*, 1994) and appears to account for a proportion of early-onset breast cancer roughly equal to BRCA1, but confers a lower risk of ovarian cancer. The remaining susceptibility to early-onset breast cancer is divided between as yet unmapped genes for familial cancer, and rarer germline mutations in genes such as TP53 (Malkin *et al.*, 1990). It has also been suggested that heterozygote carriers for defective forms of the Ataxia-Telangectasia gene are at higher risk for breast cancer (Swift *et al.*, 1976; Swift *et al.*, 1991). Late-age onset breast cancer is also often familial although the risks in relatives are not as high as those for early-onset breast cancer (Cannon-Albright *et al.*, 1994; Mettlin *et al.*, 1990). However, the percentage of such cases due to genetic susceptibility is unknown.

Breast cancer has long been recognized to be, in part, a familial disease (Anderson, 1972). Numerous investigators have examined the evidence for genetic inheritance and concluded that the data are most consistent with dominant inheritance for a major susceptibility locus or loci (Bishop and Gardner, 1980; Go *et al.*, 1983; Williams and Anderson, 1984; Bishop *et al.*, 1988; Newman *et al.*, 1988; Claus *et al.*, 1991). Early results demonstrated that at least three loci exist which convey susceptibility to breast cancer as well as other cancers. These loci are the TP53 locus on chromosome 17p (Malkin *et al.*, 1990), a 17q-linked susceptibility locus known as BRCA1 (Hall *et al.*, 1990), and one or more loci responsible for the unmapped residual. As noted above, the BRCA1 and BRCA2 genes have recently been identified. These are located on chromosomes 17q and 13q, respectively. Hall *et al.* (1990) indicated that the inherited breast cancer susceptibility in kindreds with early age onset is linked to chromosome 17q21; although subsequent studies

by this group using a more appropriate genetic model partially refuted the limitation to early onset breast cancer (Margaritte *et al.*, 1992).

The simplest model for the functional role of BRCA1 holds that alleles of BRCA1 that predispose to cancer are recessive to wild type alleles; that is, cells that contain at least one wild type BRCA1 allele are not cancerous. However, cells that contain one wild type BRCA1 allele and one predisposing allele may occasionally suffer loss of the wild type allele either by random mutation or by chromosome loss during cell division (nondisjunction). All the progeny of such a mutant cell lack the wild type function of BRCA1 and may develop into tumors. According to this model, predisposing alleles of BRCA1 are recessive, yet susceptibility to cancer is inherited in a dominant fashion: women who possess one predisposing allele (and one wild type allele) risk developing cancer, because their mammary epithelial cells may spontaneously lose the wild type BRCA1 allele. This model applies to a group of cancer susceptibility loci known as tumor suppressors or antioncogenes, a class of genes that includes the retinoblastoma gene and neurofibromatosis gene. By inference this model may also explain the BRCA1 function, as has recently been suggested (Smith *et al.*, 1992).

A second possibility is that BRCA1 predisposing alleles are truly dominant; that is, a wild type allele of BRCA1 cannot overcome the tumor forming role of the predisposing allele. Thus, a cell that carries both wild type and mutant alleles would not necessarily lose the wild type copy of BRCA1 before giving rise to malignant cells. Instead, mammary cells in predisposed individuals would undergo some other stochastic change(s) leading to cancer.

If BRCA1 predisposing alleles are recessive, the BRCA1 gene is expected to be expressed in normal mammary tissue but not functionally expressed in mammary tumors. In contrast, if BRCA1 predisposing alleles are dominant, the wild type BRCA1 gene may or may not be expressed in normal mammary tissue. However, the predisposing allele will likely be expressed in breast tumor cells.

Identification of a breast cancer susceptibility locus permits the early detection of susceptible individuals and greatly increases our ability to understand the initial steps that lead to cancer. As susceptibility loci are often altered during tumor progression, cloning these genes could also be important in the development of better diagnostic and prognostic

products, as well as better cancer therapies. Knowledge of specific mutations in the BRCA1 and BRCA2 genes, which are predisposing toward breast and/or ovarian cancer, has already led to screening of patients for these mutations. Knowledge that there is a duplication of a part of the BRCA1 gene in the chromosome, this duplication being named LBRCA1, is important for accurate testing. The high similarity between BRCA1 and LBRCA1 could lead to erroneous results or could confound the testing procedure. Knowledge of the sequence and location of LBRCA1 will enable one to avoid these problems in testing.

BRCA1 is located very near to L1A1.3B, which is a partial duplication of the 1A1.3B gene, which is located very near to LBRCA1. The L1A1.3B lies head to head within 250 base pairs of BRCA1. The overlapping of regulatory regions for the two genes may be of importance in coordinate control of the two genes. The presence of a duplication containing all or part of BRCA1 and 1A1.3B suggests that recombination events or other homology-mediated genetic rearrangements, occurring somatically or as heritable changes, could result in altered expression or inactivation of genes located within or close to the duplicated segment, including, but not limited to, the BRCA1 and 1A1.3B genes.

Finally, polymorphisms have been found in LBRCA1 and the BRCA1 promoter region. These will be useful in characterizing possible mutations in LBRCA1 and will also be useful for "diagnosing" chromosomal rearrangements involving LBRCA1. This is important because with other genes it has been shown that duplication of a segment of human DNA results in a predisposition to genomic rearrangements that are associated with disease. Such a mechanism may also occur with BRCA1 and such rearrangements may be responsible for causing cancer rather than, e.g., a missense or nonsense mutation within the gene. This mechanism may be important either to cause heritable defects or to create gene defects during the somatic growth of cells that carry no inherited defect.

SUMMARY OF THE INVENTION

The present invention relates generally to the field of human genetics. Specifically, the present invention relates to the duplication of a portion of human chromosome 17q containing the breast cancer gene BRCA1. The invention relates to the chromosomal arrangement and sequence similarities of BRCA1-L1A1.3B and LBRCA-1A1.3B. This invention further relates to LBRCA1 polymorphisms and BRCA1

promoter region polymorphisms that are useful in analyzing whether genomic rearrangements have occurred and the usefulness of this in the diagnosis and prognosis of human breast and ovarian cancer.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A is a summary of chromosomal localization of specific PCR products and restriction mapping of genomic clones for the BRCA1 promoter and its cognate. The regions of chromosome 17 contained in the rodent human hybrids ND-1 and MH-41 are indicated by solid horizontal lines (vanTuinen et al., 1987). The inferred relative locations and sizes of the genomic EcoRI fragments corresponding to those in CH40 clones 10A and 16C are shown. A conserved central EcoRI site is located between the most 5' exons of the head-to-head gene arrangements and is marked by a double-thick hash. The 16C clone was first identified as containing BRCA1-specific sequences because it showed greater hybridization to the oligonucleotide 1007-5 than the 10A clone, indicated here by heavy and light arrows, as described.

Figure 1B shows an STS analysis of YAC and P1 clones previously mapped to the BRCA1 region (Albertsen et al., 1994, Neuhausen et al., 1994) and the 10A and 16C clones described elsewhere in the specification. PCR primer combinations are as described in Table 1 or elsewhere in the disclosure.

Figure 2 is a summary of features of the BRCA1 and 1A1.3B promoter region sequences. Known exons of these two genes are indicated as solid boxes and the corresponding regions of the putative cognate genes with the same apparent intron-exon boundaries are indicated as checkered boxes. The largest segments unique to each sequence are indicated as open triangles, with the length indicated in bp. The EcoRI site marked is the central EcoRI site noted in Figure 1. The position of the BRCA1 major translation product start site is indicated in exon 2 and a similar indication is shown for a possible translation start site in the corresponding segment of LBRCA1. The approximate locations of oligonucleotide primer sequences described in Table 1 are shown. The open boxes at positions 150 and 525 on the basepair scale represent polymorphisms detected in

the BRCA1 promoter. The narrow box indicates a base difference, C or T, and the wide box indicates a trinucleotide difference, AAC or AACAAC. In U37574 (shown in the Sequence Listing as SEQ ID NO:1), these correspond to nucleotide positions 612, which is C in U37574 and 980-982 where U37574 contains a single trinucleotide element. These polymorphisms are in apparent strong linkage disequilibrium, with the C/AAC haplotype having a frequency of 0.65 on 190 tested chromosomes.

Figure 3 is a dot-plot comparison of U37574 with the sequence derived from genomic clone 10A (Genbank accession U72483 (shown in the Sequence Listing as SEQ ID NO:2)). For this comparison a window size of 15 and a match criterion of 12 were used. The positions of the known and comparable exon structures for BRCA1, LBRCA1, 1A1.3B and L1A1.3B are marked along the axes in the same format as shown in Figure 2. The significant gaps representing the largest differences between the sequences discussed in the text are indicated.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates generally to the field of human genetics. Specifically, the present invention relates to a partial duplication of a human breast cancer predisposing gene (BRCA1), and some polymorphic allelic forms which can be useful in tracking the chromosomal arrangement of BRCA1. The invention also relates to the fact that there can be mutations in the LBRCA1 and L1A1.3B promoter regions that can affect transcription of the BRCA1 and 1A1.3B genes.

Mutations in the BRCA1 gene are associated with a highly increased risk of breast or ovarian cancer development, and inheritance of defective forms of this gene may account for approximately 5% of breast cancer cases. Altered expression or effective loss of function of BRCA1 is likely to be important in sporadic breast and ovarian tumors as well (Chen et al., 1995; Holt et al., 1996). Although a complete genomic structure of BRCA1 is not yet available, a complete coding region cDNA sequence of BRCA1 has been reported (Miki et al., 1994). The cDNA structure was further elucidated in a report characterizing the promoter region of BRCA1 and describing the alternative use of exons

1a or 1b in different tissue types (Xu et al., 1995). An additional complexity of BRCA1 transcription was noted by Brown et al. (1994) who provided evidence that the 1A1.3B gene identified by Campbell et al. (1994) and believed to encode the CA125 ovarian cancer marker antigen, is transcribed in a head-to-head fashion with BRCA1, with the 5' most exons of each gene located at a distance of just 295 bp.

It is shown here that the BRCA1 and 1A1.3B promoter regions are highly similar but represent distinct copies of a genomic duplication. One copy includes the head-to-head arrangement of the 1A1.3B gene and a putative gene with 5' sequences similar to BRCA1, referred to here as LBRCA1 (for Like BRCA1). The second promoter region has a head-to-head arrangement of BRCA1 with a putative gene L1A1.3B (for Like 1A1.3B) that has a 5' structure similar to 1A1.3B. This view is supported by analysis of genomic PCR products specific for each promoter region and of genomic clones that have been propagated in recombination-deficient conditions. There is a high degree of similarity of the two sequences, but also significant differences, consistent with functional divergence since the time of the duplication event. New hypotheses regarding mechanisms of breast and ovarian cancer etiology involving the newly recognized genetic structures and putative genes are presented.

The data presented here demonstrate the existence of a direct genomic duplication that includes the BRCA1 and 1A1.3B promoters as distinct elements. The alternative forms of the duplication do not represent polymorphic variation because PCR reactions with primers specific for each distinct segment showed products of the correct size with all genomic samples ($N > 90$) and sequencing of such products showed the expected single pattern (data not shown). This finding has a wide variety of implications in part because it significantly revises a generally accepted (Szabo and King, 1995) and frequently cited aspect of BRCA1 gene structure.

The possible expression of LBRCA1 and L1A1.3B genes that include homologies to BRCA1 and 1A1.3B throughout all or part of their length could pose previously unrecognized difficulties for the development of specific antibodies and probes for precise study of gene expression and function. Conflicting and apparently inconsistent immunohistochemical data have been observed for both 1A1.3B (Campbell et al., 1994) and BRCA1 (Scully et al., 1996; Chen et al., 1996). It is also very likely that DNA and

RNA hybridization results obtained to date with genomic and cDNA probes for 1A1.3B and BRCA1 require some review. In some circumstances, evaluation of specific expression of these loci may depend on RT-PCR analysis with specific primers or the identification and verification of specific hybridization probes. More complete genomic structural characterization and transcription analysis are needed to determine the expression pattern and nature of any gene products from these loci.

Searches for mutations affecting BRCA1 transcription initiation must be carried out using primers and PCR conditions that are completely specific for amplification of the BRCA1 promoter region. The most significant published effort to screen for such mutations (Friedmann et al., 1995) relied on primers designed from what is now recognized as primarily the 1A1.3B and LBRCA1 promoter sequences. Since that strategy failed to reveal the common polymorphisms that we have detected, the primer set used could not have provided fully sensitive mutation screening coverage for the BRCA1 promoter region. This indicates the need for renewed experimental approaches to analysis of the promoter for patients with "inferred regulatory" mutations (Gayther et al., 1995). With respect to the coding regions of BRCA1, the possibility may exist that some genomic mutations assigned to BRCA1 actually reside in LBRCA1 if genomic or RT-PCR primer pairs thought to be specific for BRCA1 also amplify an identical sequence from LBRCA1. The overall sequence similarity of ~94%, observed in the promoter regions of the two genes suggests that such confusion is not likely if this degree of similarity is representative of the entire duplication. However, knowledge of the sequences of the two similar regions will be useful in the design of PCR primers needed for amplification of products specific for each region of the duplication.

The finding that L1A1.3B and not 1A1.3B is located head-to-head with BRCA1 may imply a coordinate regulation and that the putative L1A1.3B gene/transcript shares a greater functional interaction or a greater developmental and tissue-specific coordination of expression with BRCA1 than does 1A1.3B. Therefore, mutations in L1A1.3B could account for some instances of familial breast-ovarian cancer genetically linked to the BRCA1 locus, but without any known mutation yet identified in the BRCA1 gene.

A second gene involved in both sporadic and familial ovarian cancers that is distal to BRCA1 has been inferred by loss of heterozygosity (LOH) studies (Godwin et al.,

1994). Genetic linkage to 17q21 for families with a site-specific ovarian cancer susceptibility has also been established (Steichen-Gersdorf et al., 1994). The newly identified gene/promoter complex does lie distal to the BRCA1 gene at a close but not yet defined distance, suggesting that observations of specific instances of LOH in ovarian tumors that do not involve BRCA1 could involve this second locus. Some inherited site-specific ovarian cancer may also be due to mutations in the newly identified genes or promoter segments.

The presence of a duplication containing all or part of BRCA1 and 1A1.3B suggests that recombination events or other homology-mediated genetic rearrangements, occurring somatically or as heritable changes, could result in altered expression or inactivation of genes located within or close to the duplicated segment. Examples of such mechanisms include unequal exchanges resulting in the formation of chimeric genes with inappropriate expression or function (Lifton et al., 1992) and deletion or gene conversion events between highly similar gene sequences that result in non-functional arrangements (White et al., 1988). The alternative possibilities of duplication or deletion of one or more genes lying between sites of homology-mediated unequal exchange may also be involved in disease etiology. An example of this is the PMP22 gene, located in proximal 17p between two homologous 24 kb elements that are separated by 1.5 Mb (Kiyosawa and Chance, 1996). Unequal exchange between these elements can cause a duplication of PMP22, resulting in Charcot-Marie-Tooth disease Type I (Pentao et al., 1992) or a deletion that causes hereditary neuropathy with liability to pressure palsies (Chance et al., 1993).

Inversions caused by recombination between homologous 9.5 kb segments located 250-350 kb apart and in opposite orientation on the same chromosome are responsible for almost 50% of the mutations in FVIII (Naylor et al., 1993; Lakich et al., 1993; Naylor et al., 1995). It is notable that the FVIII gene in patients with hemophilia A was scrutinized for 8 years before this common mutation mechanism was detected. As was the case for FVIII, it is possible that large scale inversion, duplication or deletion mutations involving the 1A1.3B/LBRCA1 and L1A1.3B/BRCA1 segments have been missed by investigations to date. This is particularly true for the evaluation of tumor material, where appropriate DNA specimens for analyses of very long fragments are usually unavailable

and the relevant detection methods are rarely applied. Mutation studies of individual exons or even large cDNA segments would not result in identification of such changes, because at least one normal gene copy would be present in cases of a large scale rearrangement affecting one chromosome. Furthermore, PCR-based detection strategies that do not specifically anticipate changes in the order or orientation of gene segments are insensitive to such changes. Further elucidation of the complete genomic structure of the duplication described here and development of appropriate detection methods will reveal the contribution of specific long-range chromosomal rearrangements to the burden of somatic genetic events causing sporadic cancer cases as well as inherited defects.

RE-SSCP

Scanning of long PCR product fragments for DNA sequence variation was carried out by methods similar to those described previously (Liu and Sommer, 1995), except that gel-purified PCR products were uniformly labeled by 12 cycles of reamplification with the same or internal PCR primers in the presence of alpha-^{33}P-dNTPs and were then digested with a series of appropriate restriction endonucleases before application to 0.5 X MDE gels (FMC Bioproducts) for detection of SSCP and heteroduplex variants.

Genomic library screening

The LANL1701 flow-sorted chromosome 17 library (Longmire et al., 1993) was provided by L. Deaven at the Los Alamos National Laboratory. The vector, lambda CH40, grows in recA$^-$ bacteria, and the library has been propagated on the K802 recA$^-$ host to significantly reduce the possibility of intra- or inter-clone recombination events that might result in artifactual fusions. To screen this library, PCR of DNA from library subpools was used to verify the presence of appropriate clones, followed by plaque hybridization. Standard methods were used for phage clone growth, DNA extraction, restriction digestion and construction of pUC8 plasmid derivatives containing each of the EcoRI fragments of each CH40 clone for further hybridization analysis and DNA sequencing.

Oligonucleotide hybridization

Plasmid and phage clone DNAs were denatured with NaOH and applied to charged nylon filters (AMF CUNO), followed by air-drying, UV crosslinking and filter pre-washing in 0.5% SDS, 0.1 x SSC at 65°C. Oligonucleotide probes were ^{32}P-labeled with T4 kinase and hybridized to replicate filters in 6 x SSC, 5 mM EDTA pH 8.0, 0.25% non-fat dry milk for at least two hours at 37°C, followed by three successive 5 minute washes in pre-warmed 5 x SSC, 0.1% SDS at a temperature 10°C lower than the oligonucleotide T_m, as calculated by the PRIMER program. Filters were blotted dry and exposed to X-ray film for 1 to 16 days with an intensifier screen. The sequences of the oligonucleotides used are as shown in Table 1.

Sequencing and sequence analysis

Manual cycle-sequencing of clones and PCR products was carried out as described by Adams and Blakesley (1991). PCR products were purified from low melting agarose using Promega Wizards columns. Sequencher 2.0 or 3.0 software (Genecodes) was used to generate restriction maps of known sequences, for assembly of manual sequencing data and comparison of related sequences. DottyPlotter 1.0c software (BIONET, D.G. Gilbert, 1989) was used for comparison of sequences at different "stringencies" by dot plot analysis. PRIMER 0.5 (Whitehead Institute 1991, Hudson et al., 1992) was used for PCR primer analysis and design. Sequence similarity searches of Genbank and EMBL sequence databases were conducted using the BLAST suite of programs (Altschul et al., 1990) supported by the National Center for Biotechnology Information.

Chromosomal localization, STS analysis of genomic YAC and P1 clones

PCR reactions for chromosome localization and sub-localization using rodent-human hybrid DNAs were carried out as described (Barker et al., 1993) using serial dilutions of the template DNAs to allow identification of any artifactual positives due to slight contamination by cells with different chromosomal complements. DNAs were prepared by standard methods from YAC and P1 clones, obtained from the Baylor Human Genome Center and Genome Systems respectively, and similarly analyzed. STS primers for RNU2 (Genome Database), for 1A1.3B exons 12 and 19 (Campbell et al., 1994) and

TABLE 1
Sequences of Key Oligonucleotides

Primer	Sequence	Partner	Product	PCR Conditions
120.2	CCAGTACCCCAGAGCAT CA (SEQ ID NO:3)	120.3	BRCA1	(45 secs at 94°C, 60 secs at 57°C, 90 secs at 72°C) x 30
120.3	TGAACTTCCCCAAACCC TC (SEQ ID NO:4)	120.2	BRCA1	see above
214.3	TGGATGGAGAACAAGG AATC (SEQ ID NO:5)	42.2	BRCA1	(30 secs at 94°C, 60 secs at 60°C, 165 secs at 72°C) x 6 followed by (30 secs at 94°C, 60 secs at 55°C, 170 secs at 72°C) x 30
42.2	TGAACTTCTCCAAACCC TC (SEQ ID NO:6)	120.2	BRCA1	(30 secs at 94°C, 60 secs at 58°C, 90 secs at 72°C) x 30
225.1	GGGCAGAAGCAACCTGA (SEQ ID NO:7)	225.4	1A1.3B	(45 secs at 94°C, 60 secs at 61°C, 90 secs at 72°C) x 30
225.4	GGAGGGACAGAAAGAG CC (SEQ ID NO:8)	225.1	1A1.3B	see above
42.3	GGTCAGAATCGCTACCT ATTG (SEQ ID NO:9)	-		
1007.5	AGCTCGCTGAGACTTCC TG (SEQ ID NO:10)	-		
214.2	GAAGTTGTCATTTTATA AACCTTT (SEQ ID NO:11)	-		

For PCR primer pairs, the thermal cycling conditions, and the specificity of the product (BRCA1 or 1A1.3B) are as shown. Taq polymerase and standard reaction buffer were from Promega and cycling was performed in a Perkin-Elmer 480 or Techne PHC-3 thermal cycler.

for BRCA1 exon 11 (5'-AGTGATCCTCATGAGGCTTT-3' (SEQ ID NO:12) and 5'-TTAACTGTCTGTACAGGCTTGAT-3' (SEQ ID NO:13), designed using information in Genbank entry U14680 (shown in the Sequence Listing as SEQ ID NO:14)) were used for the YAC and P1 analyses, in addition to primer pairs specific for each promoter.

The present invention is described by reference to the following Examples, which are offered by way of illustration and are not intended to limit the invention in any manner. Standard techniques well known in the art or the techniques specifically described were utilized.

EXAMPLE 1

Detection of a Duplication Involving the BRCA1 Promoter Region

PCR primers 120.3 and 120.2 (Table 1) were designed for amplification of the BRCA1 promoter region from information presented by Brown et al. (1994). This published sequence is apparently an improper fusion of 1A1.3B-specific sequences with BRCA1-specific sequences, probably due to the very close similarity between the two promoter regions that caused a clone rearrangement or a false contig assignment. This initial primer pair included one (120.3) that corresponds to a region of near-identity in the 1A1.3B gene with its cognate and a second (120.2) that is BRCA1-specific. RE-SSCP analysis of the 1300 bp PCR product revealed two polymorphic sites. However, the restriction fragment patterns observed did not agree completely with those predicted from the promoter sequence of Brown et al. (1994). DNA sequencing to identify the polymorphic sites confirmed this apparent distinction. BLAST searches showed that U37574 contributed by Xu et al. (1995), was essentially identical to the segment in which the polymorphisms occurred. U37574 is a 3.8 kb genomic PstI fragment that includes the BRCA1 promoter and the alternative 5' BRCA1 exons 1a and 1b as well as upstream sequences (Xu et al., 1995).

By testing various pairs of additional primers designed from all the available sequence information, we identified primer pairs and cycling conditions (Table 1) that consistently amplified, from human genomic DNA, segments with distinct sequences that were essentially identical to portions of either the Brown et al. (1994) sequence or the

U37574 sequence. Primers 42.2 and 214.3 (Table 1) were used to amplify a 2600 base segment that extends from 1100 bp upstream of the first BRCA1 exon through BRCA1 exon 2. Sequencing of the 42.2 plus 214.3 PCR product obtained with a genomic DNA sample carrying the BRCA1 185delAG mutation in exon 2 (Simard et al., 1994; Struewing et al., 1995) revealed a simple sequence pattern heterozygous for 185delAG, demonstrating that this genomic PCR product represents BRCA1. Sequencing of portions of the upstream region of this same 42.2 plus 214.3 product were in essentially complete agreement with the U37574 sequence, confirming the correspondence of U37574 to BRCA1. In contrast, primers 225.1 plus 225.4 (Table 1) amplified a fragment with a sequence corresponding to that of Brown et al. (1994) throughout nearly its entire length, with divergence at a position close to the 225.4 primer indicating the site of the apparent artifactual fusion. Comparison of the 225.1 plus 225.4 sequence to the BRCA1-specific sequence revealed 6% non-identity of corresponding bases as well as 6 short insertion/deletion differences, demonstrating the existence of two similar but distinct genomic segments.

EXAMPLE 2

Characterization of the Location and Structure of the Duplication

The primer pairs specific for amplification of each of the distinct promoter segments were used to determine their genomic localization using human-rodent hybrid cell lines (vanTuinen et al., 1987) containing known portions of chromosome 17 (Figure 1A). In each case, a positive PCR reaction was observed with template DNA from hybrid ND-1 but not from MH-41, placing both promoter complexes in a region between hybrid breakpoints at 17q21.1 and 17q23.1, in agreement with the well-established localization of BRCA1 at 17q21 (data not shown). Analyses of DNA from additional hybrids with different chromosome 17 breakpoints, as well as DNA from hybrid MH-22, containing human chromosome 17 as its only human complement, were consistent with a unique localization at 17q21. STS analysis of genomic YAC and P1 clones included in physical maps of the BRCA1 region (Albertsen et al., 1994; Neuhausen et al., 1994) is presented in Figure 1B. The key observations are that P1 clone 746B4 contains a segment spanning

BRCA1 exon 11 and the BRCA1 promoter, but not including the 1A1.3B promoter, while YAC clone 173B7 includes the 1A1.3B promoter and 1A1.3B exons 12 and 19, but not the BRCA1 promoter. These and other data summarized in Figure 1B are most consistent with the view that the two distinct promoters are corresponding elements of a direct duplication at 17q21, with the gene loci oriented with respect to each other and the chromosome as depicted in Figure 1A.

EXAMPLE 3

Isolation and Refined Analysis of Genomic Clones Containing the Duplicated Promoter Regions

As described above, genomic clones containing the BRCA1 or 1A1.3B promoters and adjacent sequences were isolated from a rearrangement-resistant lambda library. PCR analysis of the complete library DNA indicated the presence of both promoter segments. All isolated clones hybridized strongly to the 225.1 plus 225.4 PCR product used as probe, but revealed one of two distinct EcoRI restriction patterns. Clone 10A and 5 similar isolates contained two EcoRI fragments of 7.0 and 9.2 kb. Clone 16C contained EcoRI fragments of 7.1, 2.7, 2.5, 1.5 and 0.35 kb. Plasmid DNAs containing individual EcoRI fragments of 10A and 16C were probed with oligonucleotides 42.2, 42.3, 225.4, 1007.5 and 214.2 (Figure 2, Table 1) to establish the fragment maps shown in Figure 1A. The oligonucleotide hybridization analysis also confirmed that clone 16C includes the 5' portion of BRCA1. Oligonucleotide 1007.5, corresponding to well-established BRCA1 5' cDNA sequence (Figure 2, Table 1), showed strong hybridization to the 16C 7.1 kb EcoRI fragment but weak hybridization to the 10A 9.2 kb EcoRI fragment, detectable only with long autoradiographic exposure. Sequence analysis of the termini of the 1.5 and 7.1 kb EcoRI fragments of 16C showed that these do not include any of the CH40 vector sequences and are therefore "natural" EcoRI fragments. In contrast, each of the EcoRI fragment subclones of 10 A includes an "artificial" CH40 EcoRI end, showing that the corresponding genomic EcoRI fragments must be longer than 7.0 and 9.2 kb (Figure 1A). The different EcoRI site patterns of these two clones show that the segments containing the distinct promoter complexes are not overlapping or interdigitated.

DNA sequencing of portions of the 1.5 and 7.1 kb EcoRI fragments of clone 16C as well as of the BRCA1-specific PCR products described earlier was in essentially complete agreement with the U37574 sequence. This identity was confirmed for approximately 75% of the length of U37574, including all of the region upstream of BRCA1 exon 1a, and all of BRCA1 exons 1b and 2. In contrast, the 10A sequence was found to include the 1A1.3B exon 1a and 1b sequences as reported by Brown et al. (1994). The region corresponding to the location of BRCA1 is also quite distinct.

Figure 3 shows a dot plot of the U37574 sequence vs. a corresponding segment of clone 10A. This region includes 1A1.3B exons 1a and 1b and BRCA1 exons 1a, 1b and 2 and their cognates. The strong diagonal elements of the plot indicate a high degree of sequence similarity across this entire length. There are three significant gaps in this similarity. Gap1 (Figure 3) is due to a 340 bp insertion in LBRCA1 just at the beginning of the sequence that corresponds to BRCA1 exon 1a (Figure 2). It is unclear whether this insertion may be considered part of LBRCA1 exon 1a. It does not include homology to any highly repeated human sequence.

Gap2 is due to an additional 61 bp of sequence within the segment of LBRCA1 that corresponds to BRCA1 exon 1b. As indicated in Figures 2 and 3, nearly all of the LBRCA1 "exon 1b-like" sequence and a significant part of the BRCA1 exon 1b sequence correspond to a region of homology with the Alu repeat element. The additional 61 bases that are present in the LBRCA1 gene represent that part of the Alu element that is missing from BRCA1 exon 1b. This difference strongly suggests that the Alu element at this position existed prior to the duplication event and that part of this Alu was lost in the further evolution of BRCA1 exon 1b. The finding that exon 1b is derived from an Alu element is an example of a phenomenon already described for a variety of other known genes (Makalowski et al., 1994; Baban et al., 1996). Since the Alu element is found only in primates, the proposed duplication almost certainly occurred after the genomic dispersion of this element in the primate genome. The function of exon 1b in BRCA1 is also very likely to be unique to primates.

Gap3 (Figure 3) corresponds to a region upstream of BRCA1 exon 2. At this position, LBRCA1 includes a complete Alu element in opposite orientation to the exon 1b Alu element. This Alu is missing from the BRCA1 gene. However, there are about 60

non-contiguous basepairs in BRCA1 at this position that are not present in LBRCA1. A fourth notable feature in Figure 3, the outlying diagonal element, is due to the presence of an additional Alu repeat, downstream of BRCA1 exon 2, and in the same orientation as the exon 1b element. The known sequence of the LBRCA1 segment ends within this repeat.

Differences between the 1A1.3B exons and L1A1.3B are small and only evident upon closer inspection of Figure 3. The sequences of L1A1.3B that correspond to exons 1a and 1b of 1A1.3B both reveal short deletions totaling 24 and 10 basepairs respectively. Since neither of the 1A1.3B 1a or 1b exons encodes any known translation product, the significance of these differences is not apparent.

EXAMPLE 4

Implications of DNA Structure for Expression of LBRCA1 and L1A1.3B

The fact that each of the gene complexes, L1A1.3B/BRCA1 and 1A1.3B/LBRCA1, contains one gene with well-established transcriptional activity shows that both of the newly identified promoters are active. The possibility of functional transcription of the LBRCA1 and L1A1.3B genes is supported by the overall structure and sequence similarity of the promoter complex regions (Figures 2 and 3) as well as the conservation of splicing sequences for the presumptive exons of both LBRCA1 and L1A1.3B. The BRCA1 start site and coding frame in exon 2 are not conserved in LBRCA1, however there is a potential ATG start site close to the end of the sequence that is similar to exon 2 (Figure 2).

A feature of the region of Alu similarity in BRCA exon 1b and the corresponding segment of LBRCA1 (Figure 2) that is likely to be significant for expression is the presence of an Alu-related estrogen responsive element (ERE) as defined by Norris et al. (1995), that functions as an estrogen-dependent transcription enhancer. By comparison of two functionally defined ERE elements, one of them derived from an unknown location within the 5' 50 kb of BRCA1, these authors proposed the consensus sequence $GGTCA(N)_3TGGTC(N)_9TGACC$ (SEQ ID NO:15). This sequence was found within Alu elements, aligned in reverse orientation with respect to the "sense" Alu orientation

(indicated by the arrows in Figure 2). Corresponding segments of BRCA1 exon 1b and LBRCA1 include a perfect match to this consensus, with the expected orientation relative to the Alu. Flanking sequences distinguish both of these elements from that reported by Norris et al. (1995), showing that the 5' BRCA1 region contains at least two EREs. No other consensus EREs are present in the segments shown in Figure 2.

EXAMPLE 5

Using Polymorphisms to Track BRCA1 Chromosomal Rearrangements

Several genes in the human genome that are duplicated have now been identified. These gene duplications are often a result of unequal crossing over events that have occurred during the evolutionary history of the human species. Often both elements of the duplicated segment have subsequently evolved functions that are essential for normal development and health. Events that occur during the growth of cells of a single human individual can result in unequal crossing over that reverses the effect of the evolutionary event, destroying one or both of these functions. Another possible outcome of unequal crossing over is further expansion of the duplicated region, which may also result in destruction of functional gene arrangements.

Duplications and deletions of specific genes have been associated with disease states. Unequal exchange within the PMP22 gene may result in Charcot-Marie-Tooth disease Type I if there is a duplication (Pentao et al., 1992) or it may result in hereditary neuropathy with liability to pressure palsies if there is a deletion (Chance et al., 1993). Inversions caused by recombination in the Factor VIII gene are responsible for almost 50% of the cases of hemophilia A (Naylor et al., 1993; Lakich et al., 1993; Naylor et al., 1995). The iduronate-2-sulphatase (IDS) gene is duplicated in the genome and recombination between the IDS gene and its second locus (IDS-2) is the cause of Hunter Syndrome in 13% of patients with this disease (Bondeson et al., 1995). Unequal crossing-over between 11β-hydroxylase and aldosterone synthase leading to partial duplication of both genes with the 5' regulatory region of 11β-hydroxylase fused to the coding sequence of aldosterone synthase causes glucocorticoid-remediable aldosteronism (Lifton et al., 1992).

The region containing the BRCA1 and 1A1.3B genes is partially duplicated in the human genome and this duplication enhances the chance that chromosomal rearrangement will occur via unequal crossing over between the homologous elements of the duplieated structures. This could easily result in inactivation of these genes or to other pathological gene arrangements. Determining the presence of such a mutation can be more difficult than finding a point mutation and may be missed. If screening for mutations is limited to sequencing the complete coding regions of these genes, a recombination that occurred within an intron will likely not be seen. This would be the result if the gene is sequenced by first amplifying the gene via PCR using sets of primers that amplify only the exons. The results of such screening could well show that all of the exons are present within the genome and may find no mutations such as point mutations, insertions, deletions, etc. Nevertheless, if a recombination has occurred within the gene resulting in an unequal crossing over, at least one of the two chromosomes will in fact not have an intact gene and the gene on that chromosome will be inactive.

Unequal crossing over may occur within somatic tissue or may occur in the germline. If such occurs within a cell in somatic tissue then that cell may be the start of a tumor. If the rearrangement occurs within the germline then one of the recombined chromosomes may be passed on to progeny. These descendants may receive a wild-type gene from an unaffected parent and the recombined chromosome from the affected parent. Loss of the active gene (loss of heterozygosity) within a cell in such a person will likely cause that cell to be the start of a tumor. Clearly, a person carrying a chromosome in which there has been a genetic rearrangement affecting the BRCA1 gene thereby inactivating it is at as much risk of developing breast or ovarian cancer as is a person with a point mutation or deletion or insertion within a single chromosome which is known to be associated with these cancers. Methods for detecting these rearrangements will be very useful, certainly just as useful as methods for detecting the point mutations, deletions and insertions within BRCA1 that are known to be associated with breast and ovarian cancer.

One method of tracking chromosomal rearrangements is to look at polymorphisms that occur within the genes. Several polymorphisms are now known for BRCA1. Two new polymorphisms are disclosed here. These two polymorphisms occur in the BRCA1

promoter region (nucleotides 612 and 980-982 of U37574 (SEQ ID NO:1)). Nucleotide 612 may be either C or T and nucleotides 980-982 may be either AAC (as shown in U37574) or AACAAC. These two polymorphisms show a high level of linkage disequilibrium. Because of this high linkage disequilibrium there are only two "genotypes" so far as looking at the combination of these two polymorphisms, i.e., a chromosome will have either C/AAC or it will have T/AACAAC. The C/AAC haplotype has a frequency of 0.65.

These polymorphisms may be used to track recombination within somatic tissue. If both chromosomes have the same "genotype", i.e., both are C/AAC or both are T/AACAAC then it will be uninformative to use these polymorphisms to study recombination within BRCA1. If the person is heterozygous for these polymorphisms, i.e., one chromosome contains C/AAC and the other chromosome contains T/AACAAC, then use of these polymorphisms will be perfectly informative in assaying for recombination within BRCA1. To perform such an assay, germline tissue is assayed for the presence of these two polymorphisms. If the person is heterozygous then both genotypes will be seen. Somatic tissue is also analyzed. If the somatic tissue shows only one of the two genotypes then clearly the chromosome carrying the other genotype has been deleted for the region containing at least that portion of BRCA1 containing the polymorphic site. Such a result would indicate a high probability that the suspect tissue is indeed cancerous. This would be strengthened by the knowledge that the person contains a mutation known to be associated with breast and ovarian cancer. This test confirms the loss of heterozygosity which may lead to cancer when the wild-type gene is lost. Note that if the person were homozygous then this test would not be applicable since only one genotype is present and this genotype will be seen regardless of whether there are two copies or one copy of the polymorphism present. If a person were hemizygous due to inheriting a BRCA1 gene which was partially deleted then the above assay would work in that loss of the wild-type copy of the gene would result in the presence of zero copies of the polymorphisms and this would be noted by an inability to amplify the gene region. However, one must be able to know that the person was hemizygous rather than homozygous to utilize such an assay.

Such an assay is not limited to the two polymorphisms noted above but may be used with other polymorphisms within the BRCA1/LBRCA1 region. Several polymorphisms have been published for BRCA1. Two new polymorphisms within LBRCA1 have been discovered and are presented here. One is at base 1723 of SEQ ID NO:2 and the other at base 2182 of SEQ ID NO:2. The base at both of these positions may be either G or A.

The above method is suitable for tracking recombination within somatic tissue but not in the germ line. If a person has inherited one wild-type gene and one gene with a deletion of the chromosomal region between LBRCA1 and BRCA1, the person will be hemizygous for the noted polymorphisms if the recombination has deleted them from one chromosome. This hemizygous person will appear homozygous for either the C/AAC or the T/AACAAC polymorphism if those are the polymorphisms being examined. If a person has inherited one wild-type gene and one gene with a duplication of the chromosomal region between LBRCA1 and BRCA1, the person may have three copies of the gene region containing the polymorphisms. Such a person could be either homozygous or heterozygous for the polymorphisms. Regardless of whether a person with a germline rearrangement has one copy or three copies of the gene region, if the recombination occurred within introns, simply sequencing exons will not discover this rearrangement. Nevertheless, the copy number of the gene region containing the polymorphism may be determined by methods such as quantitative PCR (see, e.g., Volkenandt et al., 1992; Filliland et al., 1990; Pastore et al., 1996) or fluorescent in situ hybridization (FISH). FISH analysis would easily discern a deletion of the region. Whereas many, possibly most, genes in the human genome are not duplicated in part or whole and genetic recombination within the gene would be expected to be quite rare, BRCA1 and its contiguous gene L1A1.3B are partially duplicated (as LBRCA1 and 1A1.3B) and this region is therefore much more likely to undergo unequal crossing over leading to gene deletion or duplication. Analysis of such is therefore more important with BRCA1 than it will be with genes which are not duplicated. The presence of the Alu repeat within the BRCA1-1A1.3B genes makes crossing over an even more likely event than for genes without such a repeat.

Analysis of the copy number of BRCA1 need not be limited to use of the polymorphic regions. Any region may be used. The importance of the analysis is that if the test results indicate the presence of only a single copy then this is equivalent to having a mutation known to be associated with breast or ovarian cancer since there is only a single (at most) wild-type copy of the gene present. If the test results indicate the presence of three copies of the gene region then again this is cause for concern because it will indicate the presence of a duplication of the gene region with the possibility that the duplication is a result of unequal crossing over which has inactivated the BRCA1 gene. If so then again there would be at most one copy of wild-type BRCA1 present. Clearly the knowledge of copy number of the BRCA1 gene is as important as knowing the presence of point mutations.

It will be appreciated that the methods and compositions of the instant invention can be incorporated in the form of a variety of embodiments, only a few of which are disclosed herein. It will be apparent to the artisan that other embodiments exist and do not depart from the spirit of the invention. Thus, the described embodiments are illustrative and should not be construed as restrictive.

LIST OF REFERENCES

Adams, S.M. and Blakesley, R. (1991). "Linear amplification DNA sequencing", *Focus* **13**:56-57.

Albertsen, H.M., Smith, S.A., Mazoyer, S., Fujimoto, E., Stevens, J., Williams, B., Rodriquez, P., Cropp, C.S., Slijepcivic, P., Carlson, M., Robertson, M., Bradley, P., Lawrence, E., Harrington, T., Sheng, Z.M., Hoopes, R., Sternberg, N., Brothman, A., Callahan, R., Ponder, B.A.J. and White, R. (1994). "A physical map and candidate genes in the BRCA1 region on chromosome 17q12-21", *Nature Genetics* 7:472-479.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990). "Basic local alignment search tool", *J. Mol. Biol.* **215**:403-410.

American Cancer Society, Cancer Facts & Figures - 1992. (American Cancer Society, Atlanta, GA).

Anderson, D.E. (1972). *J. Natl. Cancer Inst.* **48**:1029-1034.

Anderson, J.A., *et al.* (1992). *J. Otolaryngology* **21**:321.

Baban, S., Freeman, J.D. and Mager, D.L. (1996). "Transcripts from a novel human KRAB Zinc finger gene contain spliced Alu and endogenous retroviral segments", *Genomics* **33**:463-472.

Barker, D.F. and Fain, P.R. (1993). "Definition and mapping of STSs at STR and RFLP loci in Xp11-Xq22", *Genomics* **18**:712-716.

Bishop, D.T. and Gardner, E.J. (1980). In: Banbury Report 4: Cancer Incidence in Defined Populations (J. Cairns, J.L. Lyon, M. Skolnick, eds.), Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 309-408.

Bishop, D.T., *et al.* (1988). *Genet. Epidemiol.* **5**:151-169.

Bondeson, M.-L., Dahl, N., Malmgren, H., Kleijer, W.J., Tönnesen, T., Carlberg, B.-M. and Pettersson, U. (1995). "Inversion of the IDS gene resulting from recombination with IDS-related sequences is a common cause of the Hunter syndrome", *Hum. Mol. Genet.* 4:615-621.

Brown, M.A., Nicolai, H., Xu, C-F., Griffiths, B.L., Jones, K.A., Solomon, E., Hosking, L., Trowsdale, J., Black, D.M. and McFarlane, R. (1994). "Regulation of *BRCA1*", *Nature* **372**:733.

Campbell, I.G., Nicolai, H.M., Foulkes, W.D., Senger, G., Stamp, G.W., Allan, G., Boyer, C., Jones, K., Bast, R.C. Jr., Solomon, E., Trowsdale, J. and Black, D. (1994). "A novel gene encoding a B-box protein within the BRCA1 region at 17q21.1", *Hum. Mol. Genet.* **3**:589-594.

Cannon-Albright, L., *et al.* (1994). *Cancer Research* **54**:2378-2385.

Chance, P.F., Alderson, M.K., Leppig, K.A., Lensch, M.W., Matsunami, N., Smith, B., Swanson, P.D., Odelberg, S.J., Disteche, C. and Bird, T.D. (1993). "Deletion associated with hereditary neuropathy with liability to pressure palsies", *Cell* **72**:143-151.

Chen, Y., Chen, P.L., Riley, D.J., Lee, W.-H., Allred, D.C. and Osborne, C.K. (1996). "Location of BRCA1 in human breast and ovarian cancer cells: Response to Scully et al.", *Science* **272**:125-126.

Chen, Y., Chen, C.-F., Riley, D. J., Allred, D. C., Chen, P.-L., Von Hoff, D., Osborne, C. K. and Lee, W.-H. (1995). "Aberrant subcellular localization of BRCA1 in breast cancer", *Science* **270**: 789-791.

Claus, E., *et al.* (1991). *Am. J. Hum. Genet.* **48**:232-242.

Dunn I.S. and Blattner, F.R. (1987). "Charons 36-40: multi-enzyme, high capacity, recombination deficient replacement vectors with polylinkers and polystuffers", *Nucl. Acids Res.* **15**:2677-2698.

Easton, D., *et al.* (1993). *Am. J. Hum. Genet.* **52**:678-701.

Friedmann, L.S., Szabo, C.I., Ostermeyer, E.A., Dowd, P., Butler, L., Park, T., Lee, M.K., Goode, E.L., Rowell, S.E. and King, M.-C. (1995). "Novel inherited mutations and variable expressivity of BRCA1 alleles, including the founder mutation 185delAG in Ashkenazi Jewish families", *Am. J. Hum. Genet.* **57**:1284-1297.

Futreal, P.A., *et al.* (1994) "*BRCA1* Mutations in Primary Breast and Ovarian Carcinomas", *Science* **266**:120-122.

Gayther, S.A., Warren, W., Mazoyer, S., Russell, P.A., Harrington, P.A., Chiano, M., Seal, S., Hamoudi, R., van Rensburg, E.J., Dunning, A.M., Love, R., Evans, G., Easton, D., Clayton, D., Stratton, M.R. and Ponder, B.A.J. (1995). "Germline mutations of the BRCA1 gene in breast and ovarian cancer families provide evidence for a genotype-phenotype correlation", *Nature Genetics* **11**:428-433.

Gilliland, G., Perrin, S., Blanchard, K and Bunn, H.F. (1990). "Analysis of cytokine mRNA and DNA: Detection and quantitation by competitive polymerase chain reaction", *Proc. Natl. Acad. Sci. USA* **87**:2725-2729.

Go, R.C.P., *et al.* (1983). *J. Natl. Cancer Inst.* **71**:455-461.

Godwin, A.K., Vanderveer, L., Schultz, D.C., Lynch, H.T., Altomare, D.A., Buetow, K.H., Daly, M., Getts, L.A., Masny, A., Rosenblum, N., Hogan, M., Ozols, R.F. and Hamilton, T.C. (1994). "A common region of deletion on chromosome 17q in both

sporadic and familial epithelial ovarian tumors distal to BRCA1", *Am. J. Hum. Genet.* **55**:666-677.

Hall, J.M., *et al.* (1990). *Science* **250**:1684-1689.

Holt, J.T., Thompson, M.E., Szabo, C., Robinson-Benion, C., Arteaga, C.L., King, M.-C. and Jensen, R.A. (1996). "Growth retardation and tumour inhibition by BRCA1", *Nature Genet.* **12**:298-302.

Hudson, T.J., Engelstein, M., Lee, M.K., Ho, E.C., Rubenfield, M.J., Adams, C.P., Housman, D.E. and Dracopoli, N.C. (1992). "Isolation and chromosomal assignment of 100 highly informative human simple sequence repeat polymorphisms", *Genomics* **13**:622-629.

Kamb, A., *et al.* (1994). *Science* **264**:436-440.

Kiyosawa, H. and Chance, P.F. (1996). "Primate origin of the CMT1A-REP repeat and analysis of a putative transposon-associated recombinational hotspot", *Hum. Mol. Genet.* **5**:745-753.

Knudson, A.G. (1993). *Nature Genet.* **5**:103.

Lakich, D., Kazazian, H.H. Jr., Antonarakis, S.E. and Gitschier, J. (1993). "Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A", *Nature Genetics* **5**:236-241.

Lifton, R.P., Dluhy, R.G., Powers, M., Rich, G.M., Cook, S., Ulick, S. and Lalouel, J.-M. (1992). "A chimaeric 11 beta-hydroxylase aldosterone synthase gene causes glucocorticoid-remediable aldosteronism and human hypertension", *Nature* **355**:262-265.

Liu, Q. and Sommer, S.S. (1995). "Restriction endonuclease fingerprinting (REF): A sensitive method for screening mutations in long, contiguous segments of DNA", *BioTechniques* **18**:470-477.

Longmire, J.L., Brown, N.C., Meincke, L.J., Campbell, M.L., Albright, K.L., Fawcett, J.J., Campbell, E.W., Moyzis, R.K., Hildebrand, C.E., Evans, G.A. and Deaven, L.L. (1993). "Construction and characterization of partial digest DNA libraries made from flow-sorted human chromosome 16", *GATA* **10**:69-76.

Makalowski, W., Mitchell, G.A. and Labuda, D. (1994). "Alu sequences in the coding regions of mRNA: a source of protein variability", *Trends Genet.* **10**:188-193.

Malkin, D., *et al.* (1990). *Science* **250**:1233-1238.

Margaritte, *et al.* (1992). *Am. J. Hum. Genet.* **50**:1231-1234.

Mettlin, C., *et al.* (1990). *American Journal of Epidemiology* **131**:973-983.

Miki, Y., *et al.* (1994). "A Strong Candidate for the Breast and Ovarian Cancer Susceptibility Gene *BRCA1*", *Science* **266**:66-71.

Narod, S.A., *et al.* (1991). *The Lancet* **338**:82-83.

Naylor, J., Brinke, A., Hassock, S., Green, P.M. and Giannelli, F. (1993). "Characteristic mRNA abnormality found in half the patients with severe haemophilia A is due to large DNA inversions", *Hum. Mol. Genet.* **2**:1773-1778.

Naylor, J.A., Buck, D., Green, P., Williamson, H., Bentley, D. and Giannelli, F. (1995). "Investigation of the factor VIII intron 22 repeated region (int22h) and the associated inversion junctions", *Hum. Mol. Genet.* **4**:1217-1224.

Neuhausen, S.L., Swensen, J., Miki, Y., Liu, Q., Tavtigian, S., Shattuck-Eidens, D., Kamb, A., Hobbs, M.R., Gingrich, J., Shizuya, H., Kim, U.-J., Cochran, C., Futreal, P.A., Wiseman, R.W., Lynch, H.T., Tonin, P., Narod, S., Cannon-Albright, L., Skolnick, M.H. and Goldgar, D.E. (1994). "A P1-based physical map of the region from D17S776 to D17S78 containing the breast cancer susceptibility gene BRCA1", *Hum. Mol. Genet.* **3**:1919-1926.

Newman, B., *et al.* (1988). *Proc. Natl. Acad. Sci. USA* **85**:3044-3048.

Norris, J., Fan, D., Aleman, C., Marks, J.R., Futreal, P.A., Wiseman, R.W., Iglehart, J.D., Deininger, P.L. and McDonnell, D.P. (1995). "Identification of a new subclass of Alu DNA repeats which can function as estrogen-receptor dependent transcriptional enhancers", *J. Biol. Chem.* **270**:22777-22782.

Pastore, L., Caporaso, M.G., Frisso, G., Orsini, A., Santoro, L., Sacchetti, L. and Salvatore, F. (1996). "A quantitative polymerase chain reaction (PCR) assay completely discriminates between Duchenne and Becker muscular dystrophy deletion carriers and normal females", *Molecular and Cellular Probes* **10**:129-137.

Pentao, L., Wise, C.A., Chinault, A.C., Patel, P.I. and Lupski, J.R. (1992). "Charcot-Marie-Tooth type 1A duplication appears to arise from recombination at repeat sequences flanking the 1.5 Mb monomer unit", *Nature Genet.* **2**:292-300.

Scully, R., Ganesan, S., Brown, M., DeCaprio, J.A., Cannistra, S.A., Feunteun, J., Schnitt, S. and Livingston, D.M. (1996). "Location of BRCA1 in human breast and ovarian cancer cells", *Science* **272**:123-125.

Simard, J., Tonin, P., Durocher, F., Morgan, K., Rommens, J., Gingras, S., Samson, C., Leblanc, J.-F., Belanger, C., Dion, F., Liu, Q., Skolnick, M., Goldgar, D., Shattuck-Eidens, D., Labrie, F. and Narod, S. A. (1994). "Common origins of BRCA1 mutations in Canadian breast and ovarian cancer families", *Nature Genet.* **8**:392-398.

Smith, S.A., *et al.* (1992). *Nature Genetics* 2:128-131.

Steichen-Gersdorf, E., Gallion, H.H., Ford, D., Girodet, C., Easton, D.F., DiCioccio, R.A., Evans, G., Ponder, M.A., Pye, C., Mazoyer, S., Noguchi, T., Karengueven, F., Sobol, H., Hardouin, A., Bignon, Y.-J., Piver, M.S., Smith, S.A. and Ponder, B.A.J. (1994). "Familial site-specific ovarian cancer is linked to BRCA1 on 17q12-21", *Am. J. Hum. Genet.* 55:870-875.

Struewing, J.P., Abeliovich, D., Peretz, T., Avishai, N., Kaback, M.M., Collins, F.S. and Brody, L.C. (1995). "The carrier frequency of the BRCA1 del185AG mutation is approximately 1 percent in Ashkenazi Jewish individuals", *Nature Genetics* **11**:198-200.

Swift, M., *et al.* (1976). *Cancer Res.* **36**:209-215.

Swift, M., *et al.* (1991). *N. Engl. J. Med.* **325**:1831-1836.

Szabo, C. and King, M.-C. (1995). "Inherited breast and ovarian cancer", *Hum. Mol. Genet.* 4:1811-1817.

Tavtigian, S.V., *et al.* (1996). "The complete *BRCA2* gene and mutations in chromosome 13q-linked kindreds", *Nat. Genet.* **12**:333-337.

vanTuinen, P., Rich, D.C., Summers, K.M. and Ledbetter, D. (1987). "Regional mapping panel for human chromosome 17: Application to Neurofibromatosis Type 1", *Genomics* 1:374-381.

Volkenandt, M., Dicker, A.P., Banerjee, D., Fanin, R., Schweitzer, B., Horikoshi, T., Danenberg, K., Danenberg, P. and Bertino, J.R. (1992). "Quantitation of Gene Copy Number and mRNA Using the Polymerase Chain Reaction", *Proc. Soc. Exp. Biol. Med.* **200**:1-6.

White, P.C., Vitek, A., Dupont, B. and New, M.I. (1988). "Characterization of frequent deletions causing steroid 21-hydroxylase deficiency", *Proc. Natl. Acad. Sci. USA* **85**:4436-4440.

Williams and Anderson (1984). *Genet. Epidemiol.* **1**:7-20.

Wooster, R., *et al.* (1994). *Science* **265**:2088.

Wooster, R., *et al.* (1995). "Identification of the breast cancer susceptibility gene *BRCA2*", *Nature* **378**:789-792.

Xu, C.-F.,-Brown, M.A., Chamber, J.A., Griffiths, B., Nicolai, H. and Solomon, E. (1995). "Distinct transcription start sites generate two forms of BRCA1 mRNA", *Hum. Mol. Genet.* 4:2259-2264.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: University of Utah Research Foundation
Barker, David F.
Liu, Xudong

(ii) TITLE OF INVENTION: The BRCA1 and 1A1.3B Promoters are Parallel Elements of a Genomic Duplication at 17q21

(iii) NUMBER OF SEQUENCES: 15

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Rothwell, Figg, Ernst & Kurz, PC
(B) STREET: 555 Thirteenth Street, N.W., Suite E-701
(C) CITY: Washington
(D) STATE: DC
(E) COUNTRY: USA
(F) ZIP: 20004

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: WO
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 60/031,102
(B) FILING DATE: 26-NOV-1997
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Saxe, Stephen A.
(B) REGISTRATION NUMBER: 38,609
(C) REFERENCE/DOCKET NUMBER: 2323-114PCT

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 202-624-1589
(B) TELEFAX: 202-783-6031

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3798 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CTGCTGGNCC GGGTGCTAGG NCCCTGACTG CCCGGGGCCG GGGGTGCGGG GCCCGCTGAG 60

CCCGCGCCCA CCTGGAACTC GCGCTGGCTG GCGAGCGCTG CGCGCAGNCC CAGTTCCCAC 120

ACCCGCCTCT CCCTCCACAC TTCCCCGCAA GCAGAGGGAG CCGGCTCTGG CTTCGGCCAG 180

CCCAGAGAGG GGCCCCCACA GCGCAGTGGC GGGCTGAAGG GCTCCTCCAG CACGGNCAGA 240

ATGGACGCCA AGGNCGAGGA GGCGCCGAGA GCGAGCGAGG GCTGCTAGCA CGTTGTCACC 300

TCGCATTCTG AACCACAGAC TCTCCAACTC TCCGGNGCTT TTCGCCCACT CGGTCCCTCA 360

GAACACGAAG GGCTCTCTCA TCCTGTCACT AAAACGATTA GCTGTCCGGA GACACGGAAA 420

AAGTCGCCCC TCTTCTTTGC AGGATTCCTC CCTTGAACTT CTCCAAACCC TCTTAGTGTG 480

ACGTGACCCC ACCCCTAGCT AACCCAGGCT GCTTCCTTAC CAGCTTCCCG CCCCCTGGGG 540

AGGCGGCAAT GCAAAGACCG TCCGCTGCCA GCTCTGCCGC TATCTCTGTG GGGTGAATCT 600

AACATGGCGG ACAAAGACAG TAACTAGTCC CGTTTCTCCG CGTTTTCGCC AAGAAGATTG 660

GCTCTTACCA CTTGTCCCTC AAAACGACCA CCCCATTGAC TGGTGGCGAT TGCGTCGACG 720

GAGACGGGGC AAAAGCAAGC TGAACCCGAA AAATAACAAA CACTGGGGCT GAGGGGTGGA 780

ACTACGAGTG CGCAGACATG GGCCAGAGCG CATTTCCCCT GCCCCAGGCA AATTCGGCGC 840

TCACTGCGTC CCCGCAGGCC ACTGACCTTA CAAGACTACT TGCCCCAGAC TCCTGGGGCT 900

GGATGGGAAT TGTAGTCTCC CTAAAGAGTT GTACGTATCT TTTTAAGGCC TAGTTTCTGC 960

TTTCNAAATA CGAAAACATA ACACTCCAGT CCATAACTGT TGACAAGTAC AAGCGCGCAC 1020

AGGTCTCCAA TCTATCCACT GGATTTCCGT GAGAATTGTG CCCGCTCTGG TATTGGATGT 1080

TCCTCTCCAT AAGACTACAG TTTCTAAGGA ACACTGTGGC GAAGACCTTT CATTCCGCAA 1140

CGCATGCTGG AAATAATTAT TTCCCTCCAC CCCCCCAACA ATCCTTATTA CTTATATTTA 1200

CCGAAACTGG AGACCTCCAT TAGGGCGGAA AGAGTGGGGG ATTGGGACCT CTTCTTACGA 1260

CTGCTTTGGA CAATAGGTAG CGATTCTGAC CTTCGTACAG CAATTACTGT GATGCAATAA 1320

GCCGCAACTG GAAGAGTAGA GGCTAGAGGG CAGGCACTTT ATGGCAAACT CAGGTAGAAT 1380

TCTTCCTCTT CCGTCTCTTT CCTTTTACGT CATCCGGGGG CAGACTGGGT GGCCAATCCA 1440

GAGCCCCGAG AGACGCTTGG CTCTTTCTGT CCCTCCCATC CTCTGATTGT ACCTTGATTT 1500

CGTATTCTGA GAGGCTGCTG CTTAGCGGTA GCCCCTTGGT TTCCGTGGCA ACGGAAAAGC 1560

GCGGGAATTA CAGATAAATT AAAACTGCGA CTGCGCGGCG TGAGCTCGCT GAGACTTCCT 1620

GGACGGGGGA CAGGCTGTGG GGTTTCTCAG ATAACTGGGC CCCTGCGCTC AGGAGGCCTT 1680

CACCCTCTGC TCTGGGTAAA GGTAGTAGAG TCCCGGGAAA GGGACAGGGG GCCCAAGTGA 1740

TGCTCTGGGG TACTGGCGTG GGAGAGTGGA TTTCCGAAGC TGACAGATGG GTATTCTTTG 1800

ACGGGGGGTA GGGGCGGAAC CTGAGAGGCG TAAGGCGTTG TGAACCCTGG GGAGGGGGGC 1860

AGTTTGTAGG TCGCGAGGGA AGCGCTGAGG ATCAGGAAGG GGGCACTGAG TGTCCGTGGG 1920

GGAATCCTCG TGATAGGAAC TGGAATATGC CTTGAGGGGG ACACTATGTC TTTAAAAACG 1980

TCGGCTGGTC ATGAGGTCAG GAGTTCCAGA CCAGCCTGAC CAACGTNGGT GAAACTCCGT 2040

CTCTACTAAA AATACAAAAA TTAGCCGGGC GTGGTGCCGC TCCAGCTACT CAGGAGGCTG 2100

AGGCAGGAGA ATCGCTAGAA CCCGGGAGGC GGAGGTTGCA GTGAGCCGAG ATCGCGCCAT 2160

TGCACTCCAG CCTGGGCGAC AGAGCGAGAC TGTCTCAAAA CAAAACAAAA CAAAACAAAA 2220

CAAAAAACAC CGGCTGGTAT GTATGAGAGG ATGGGACCTT GTGGAAGAAG AGGTGCCAGG 2280

AATATGTCTG GGAAGGGGAG GAGACAGGAT TTTGTGGGAG GGAGAACTTA AGAACTGGAT 2340

CCATTTGCGC CATTGAGAAA GCGCAAGAGG GAAGTAGAGG AGCGTCAGTA GTAACAGATG 2400

CTGCCGGCAG GGATGTGCTT GAGGAGGATC CAGAGATGAG AGCAGGTCAC TGGGAAAGGT 2460

TAGGGGCGGG GAGGCCTTGA TTGGTGTTGG TTTGGTCGTT GTTGATTTTG GTTTTATGCA 2520

AGGGAAAGAA AACAACCAGA AACATTGGAG AAAGCTAAGG CTACCACCAC CTACCCGGTC 2580

AGTCACTCCT CTGTAGCTTT CTCTTTCTTG GAGAAAGGAA AAGACCCAAG GGGTTGGCAG 2640

CAATATGTGA AAAAATTCAG AATTTATGTT GTCTAATTAC AAAAAGCAAC TTCTAGAATC 2700

TTTAAAAATA TAGGACGTTG TCATTAGTTC TTTGGTTTGT ATTATTCTAA AACCTTCCAA 2760

ATCTTAAATT TACTTTATTT TAAAATGATA AAATGAAGTT GTCATTTTAT AAACCTTTTA 2820

AAAAGATATA TATATATGTT TTTCTAATGT GTTAAAGTTC ATTGGAACAG AAAGAAATGG 2880

ATTTATCTGC TCTTCGCGTT GAAGAAGTAC AAAATGTCAT TAATGCTATG CAGAAAATCT 2940

TAGAGTGTCC CATCTGGTAA GTCAGCACAA GAGTGTATTA ATTTGGGATT CCTATGATTA 3000

TCTCCTATGC AAATGAACAG AATTGACCTT ACATACTAGG GAAGAAAAGA CATGTCTAGT 3060

AAGATTAGGC TATTGTAATT GCTGATTTTC TTAACTGAAG AACTTTAAAA ATATAGAAAA 3120

TGATTCCTTG TTCTCCATCC ACTCTGCCTC TCCCACTCCT CTCCTTTTCA ACACAAATCC 3180

TGTGGTCCGG GAAAGACAGG GACTCTGTCT TGATTGGTTC TGCACTGGGG CAGGAATCTA 3240

GTTTAGATTA ACTGGCATTT TGGCTTTTCT TCCAGCTCTA AAACAAGCTC CATCACTTGA 3300

AATGGCAAAA TAAAATCATG GATGAGGCCG AGGGCGGTGG CTTATGCCTG TAATCCCAGC 3360

ACTTTGGGAG GCCAAGGTGG TAGGATCACG AGGTCAGGAG ATCGAGACCA TCCTGGCCAA 3420

CATGGTGAAA CCCCCTCTCC ACTAAAAATA CAAAAATTAG CTGGGCGTAG TGGCATGTGC 3480

CTGTAATCCC AGCTACTCAG GAGGCTGAGG CAGGAGAATC ACTTGAACCA GGAGGCAGAT 3540

GTTGCTGTGA GCCAATATGG CACCACTGAA CTCCAGCGAC AGAGCTAAAC TCCATCTCAA 3600

AAAAAAAAAA AAAAAAAAAN AAACATGGAT GATCGGTGTC GTTGAGAGGA TAGGTATTTG 3660

GAAGAACCTT TGTTTGAAAC TGGCTCTGTA CATACAATGA AATTACATAC TTATTTACAT 3720

ACAATGAAAT GCAGAGGTTT TTTTTTTATA TAGGATCTCT GTCGAGAGGC TGGAGTGCAG 3780

TGGTGCTATC ACAGCTCA 3798

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3800 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GGGAGTCGTG CCTTCCATCA GTAGAAGCCG GATGTTCTGA CCCACAGACT CTCCAACTCT 60

CCGGCGCTTC TCGCCAACTC GGTCCCTCTG AACATGAAGG GCTCTCTCAT CCTGTCACTA 120

AAAAGATTAG CTGTCCCGAA ACACGGAAAA AGTCGCCCCT CTTCTTTGCA GGATTCCTCC 180

CTTGAACTTC CCCAAACCCT CTTAGCGTGA CGTGACCCCA CCCCTAGGTA ACCGCAGCTG 240

CTTCCTTACC AGCTTCCCGC CCCCGGGGGG CGCCTGCCGG AGGCCAATGC AAGGACCGTC 300

CGCTACCGGC TCTGCCGCTA TCCCTGTGGG GTGAATCTAA CATGGCGGAT AAAGACAGTA 360

ACTAGTCCCC TGTTTCTCCG AGTGTTCGCC AAGATGATTG GCTCTCACCA CTTGTCCCTC 420

AAAACGACCA CGCCATTGAT TGGTGGAGAT TGCGTCGATG GGGCGGGGCA GAAGCAACCT 480

GAACCCGAAC AACAATAACA AACATTGAGG CTGAGGGGCG GAACTAGGAG TGCGCAGATG 540

TGGGCCAGAG CGGATTTCCC CTTCCCCAGG CAAATTCGGC GCCCACTGCG TCCCCGCAGG 600

CCACTGACCT TAGAGGACTA CTTGCCCGAG ACTCGTGGGG CTGGATGGGA ATCGTAGTCT 660

TCCTAGGAGT TGTAGGTATC TTTTTTTGGC CTAGTCTCTG CTCTCAAGAT AGGAGAACAT 720

AACAACACTC CAATCCATTA CTGTTGACAT GTATAAGCCC GCGGAGGTCT CCAATCTATC 780

CACTGGATTT CCGTGAGAAT TGTGCCCGCT TTGGTATTGG ATGTTCCTCT CCATAAGACT 840

ACAGTTTCCA AGGAACAGTG TGGCCAAGGC CTTTCGTTCC GCAATGCATG TTGGAAATAG 900

TAGTTCTTTC CCTCCACCTC CCAACAATCC TTTTATTTAC CTAAACTGGA GACCTCCATT 960

AGGGCGGAAA GAGTGGGGTA ATGGGACCTC TTCTTAAGAC TGCTTTGGAC ACTATCTTAC 1020

GCTGATATTC AGGCCTCAGG TGGCGATTCT GACCTTGGTA CAGCAATTAC TGTGACGTAA 1080

TAAGCCGCAA CTGGAAGCGT AGAGGCGAGA GGGCGGGCGC TTTACGGCGA ACTCAGGTAG 1140

AATTCTTCCT TTTCCGTCTC TTTCTTTTTA TGTCACCAGG GGAGGACTGG GTGGCCAACC 1200

CAGAGCCCCG AGAGATGCTA GGCTCTTTCT GTCCCGCCCT TCCTCTGACT GTGTCTTGAT 1260

TTCCTATTCT GAGAGGCTAT TGCTCAGCGG TTTCCGTGGC AACAGTAAAG CGTGGGAATT 1320

ACAGATAAAT TAAAACTGTG GAACCCCTTT CCTCGGCTGC CGCCAAGGTG TTCGGTCCTT 1380

CCGAGGAAGC TAAGGCCGCG TTGGGGTGAG ACCCTCACTT CATCCGGTGA GTAGCACCGC 1440

GTCCGGCAGC CCCAGCCCCA CACTCGCCCG CGCTATGGCC TCCGTCTCCC AGCTTGCCTG 1500

CATCTACTCT GCCCTCATTC TGCAGGACTA TGAGGTGACC TTTACGGAGG ATAAGATCAA 1560

TGCCCTTATT AAAGCAGCCA GTGTAAATAT TGAAACTTTT TGGCCTGGCT TGTTTGCAAA 1620

GGTCCTGGCC AACGTCAACA TTGGGAGCCA CATCTGCAAT GTAGAGGGGG GGAAAAAAAC 1680

GTGACTGCGC GTCGTGAGCT CGCTGAGACG TTCTGGACGG GGGACAGGCC GTGGGGTTTC 1740

TCAGATAACT GGGCCCCTGG GCTCAGGAGG CCTGCACCCT CTGCTCTGGG TTAAGGTAGA 1800

AGAGCCCCGG GAAAGGGACA GGGGCCCAAG GGATGCTCCG GGGGACGGGC GGGGGAAAGT 1860

GAATTTCCGA AGCTAGGCAG ATGGGTATTC TTATGCGAGG GGCGGGGGCG GAACCTGAGA 1920

GGCATAAGGC GTTGTGAACC CCCCGGGGAA GGGGGCAGTT TGTAGGTCTC GAGGGAAGCA 1980

CTAAGGATCA GGTTGGGGGC ACAGTGTGTC CGAGGAGGAA TCCTCCTGAT AGGAACTGGA 2040

ATGTGCCTTG AAGGGGACAC CATGTGTATA AGAACATCAG CTGGTCGCCG GGGATGGTGG 2100

CTTACGCCTG TATTCCTAGC ACTTTGGGAG GCCAAGGCGG ATGGATCACG AGGTCAGGAG 2160

TTCGAGACCA GCCTGACCAT CGTGGTGAAA CCCCGTCTCT ACTAAAAATA CAAAAATTAG 2220

CCGGGCGTGG TGGCGCGCGC CAGCTACTCA GGAGCTGAGG CAGGAGAATC GCTTGAACCC 2280

AGGAGGCGGA GGTTGCAGTG AGCCGAGATC GCGCCATTGC ACTCCAGCCT GGGTGGCAGA 2340

ACGACACTCC GTCTCAAAAA CAAACAAAGA AATAAACACC GGCTGGTATA TATGAGAAGA 2400

TGGGCCCTTG CGGAAGAAGA AGTGCCAGGA ATATGTCTGG GAAGGGGAGG AGACAGGATT 2460

TTGTGGGAGG GAGAACTTAA GAACTGGATC CATTTGTGCT ATTGAGAAAG CGCAAGAGGG 2520

AAGTAGAGGA GCGTCAGTAG TAACAGATGC TGCCGGCAGG GATGTGCTTG AGGGGGATCC 2580

TGAGATGAGA GTGGGTCGCT GGGAAAGGCT AGGGGCAGGG AGGCCTTGAT TGGTGTTGGT 2640

TTGGTCGTTG TTGATTTTGG TTTTATGCAA GAAAAAGAAA ACAGCCAGAA GCATTGGAGA 2700

AAGCTCACCA CTTACCCGGT CAGTCACTCC CCTGTAGCTT TCTCTTTCTT GGAGAAAGGA 2760

AAAGACCCAA AGGGTTGGAA GCAATATGTG AAAAAATACA GAATTTATAT TGTCTAATTA 2820

CAAAAAGCAA CTTCTAGAAC CTTTAAAGGA TTTTGTATTA TTCTAAAACC TTCCAAATCT 2880

TAAATTTACC TTATTTTATT TTATTTATTT NTGAGACGGA GCTTCGCTCT TGTTGCCCAG 2940

GCTGGAGTGT AATCGGCGTG ATTTGGGCTC ACCGCAACCT CTGACTCGTG GGTTCAAGCG 3000

ATTCTCCTGC CTCAGCTCCC GAGTAGCTGG GATTACACGC ATGCACCACC ATGCCTGGCT 3060

CATTTTTTTG TATTTTTAGT AGAAACGAGG TTTCTCCGTA TTGGTCAGGC TGGTCTTGAA 3120

CTCCCGACCT CAGGTCATCC GCCCGCCTCG GCCTCCCTAA GTGCTGTGAT TGACAGGCGT 3180

GAGCCACCGA CGCCCAGCCC AATTTACCTT ATTTTAAAAT GATAAAATGA AGTTGTCATT 3240

TTTCTAAACC TTTTTAAAAG ATACATGTTT TTCTAATGTG TTAAAGTTCA TTGGAACAGA 3300

AAGAGATAGA TTTATCTGCT GTTTGCGTTG AAGAAGTACA AAATGTCCTT AATGCTATGC 3360

AGAAAATCTT ACAGTGTCCA ATCTGGTAAG TCACCAGAAG AGGGTATTAA TTTGGGATTC 3420

CTATATGATT ATCTCCTATG CAAATGAACA GAATTGACCT TACATAGAAG GGAGGAAAAG 3480

ACATGTCTAA TAAGATTAGG CTATTGTAAT TGCTGATTTT CTTAACTGAA GAACTTTAAA 3540

AGTATAGAAA ATGAATCCTT GTTCTCCATC CACTCTGCCT CTCCCACTCC TCTCCTCTTC 3600

AACACAAATC CTGTGGTCCC TGAAAGACAG GGACCCTGTC TTGATTGGTT CTGCACTGGG 3660

GCAGGAATCT AGTTTAGATT AACTGGCATT TTGGTTTTNT TCTAGCTCTA AAACCAGCTC 3720

CATCACTTGA AATGGCAAAA TAAATCATGA ATGAGGCCGG GGGCTGTGGC TCACACCTGT 3780

AATCCCAGCA CTCTGGGGGG 3800

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "PCR primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CCAGTACCCC AGAGCATCA 19

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "PCR primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

TGAACTTCCC CAAACCCTC 19

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "PCR primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TGGATGGAGA ACAAGGAATC 20

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "PCR primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TGAACTTCTC CAAACCCTC 19

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "PCR primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GGGCAGAAGC AACCTGA 17

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "PCR primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GGAGGGACAG AAAGAGCC 18

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "PCR primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GGTCAGAATC GCTACCTATT G 21

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "PCR primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

AGCTCGCTGA GACTTCCTG 19

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "PCR primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GAAGTTGTCA TTTTATAAAC CTTT 24

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "PCR primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

AGTGATCCTC ATGAGGCTTT 20

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "PCR primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TTAACTGTCT GTACAGGCTT GAT 23

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5711 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

AGCTCGCTGA GACTTCCTGG ACCCCGCACC AGGCTGTGGG GTTTCTCAGA TAACTGGGCC 60

CCTGCGCTCA GGAGGCCTTC ACCCTCTGCT CTGGGTAAAG TTCATTGGAA CAGAAAGAAA 120

TGGATTTATC TGCTCTTCGC GTTGAAGAAG TACAAAATGT CATTAATGCT ATGCAGAAAA 180

TCTTAGAGTG TCCCATCTGT CTGGAGTTGA TCAAGGAACC TGTCTCCACA AAGTGTGACC 240

ACATATTTTG CAAATTTTGC ATGCTGAAAC TTCTCAACCA GAAGAAAGGG CCTTCACAGT 300

GTCCTTTATG TAAGAATGAT ATAACCAAAA GGAGCCTACA AGAAAGTACG AGATTTAGTC 360

AACTTGTTGA AGAGCTATTG AAAATCATTT GTGCTTTTCA GCTTGACACA GGTTTGGAGT 420

ATGCAAACAG CTATAATTTT GCAAAAAAGG AAAATAACTC TCCTGAACAT CTAAAAGATG 480

AAGTTTCTAT CATCCAAAGT ATGGGCTACA GAAACCGTGC CAAAAGACTT CTACAGAGTG 540

AACCCGAAAA TCCTTCCTTG CAGGAAACCA GTCTCAGTGT CCAACTCTCT AACCTTGGAA 600

CTGTGAGAAC TCTGAGGACA AAGCAGCGGA TACAACCTCA AAAGACGTCT GTCTACATTG 660

AATTGGGATC TGATTCTTCT GAAGATACCG TTAATAAGGC AACTTATTGC AGTGTGGGAG 720

ATCAAGAATT GTTACAAATC ACCCCTCAAG GAACCAGGGA TGAAATCAGT TTGGATTCTG 780

CAAAAAAGGC TGCTTGTGAA TTTTCTGAGA CGGATGTAAC AAATACTGAA CATCATCAAC 840

CCAGTAATAA TGATTTGAAC ACCACTGAGA AGCGTGCAGC TGAGAGGCAT CCAGAAAAGT 900

ATCAGGGTAG TTCTGTTTCA AACTTGCATG TGGAGCCATG TGGCACAAAT ACTCATGCCA 960

GCTCATTACA GCATGAGAAC AGCAGTTTAT TACTCACTAA AGACAGAATG AATGTAGAAA 1020

AGGCTGAATT CTGTAATAAA AGCAAACAGC CTGGCTTAGC AAGGAGCCAA CATAACAGAT 1080

GGGCTGGAAG TAAGGAAACA TGTAATGATA GGCGGACTCC CAGCACAGAA AAAAAGGTAG 1140

ATCTGAATGC TGATCCCCTG TGTGAGAGAA AAGAATGGAA TAAGCAGAAA CTGCCATGCT 1200

CAGAGAATCC TAGAGATACT GAAGATGTTC CTTGGATAAC ACTAAATAGC AGCATTCAGA 1260

AAGTTAATGA GTGGTTTTCC AGAAGTGATG AACTGTTAGG TTCTGATGAC TCACATGATG 1320

GGGAGTCTGA ATCAAATGCC AAAGTAGCTG ATGTATTGGA CGTTCTAAAT GAGGTAGATG 1380

AATATTCTGG TTCTTCAGAG AAAATAGACT TACTGGCCAG TGATCCTCAT GAGGCTTTAA 1440

TATGTAAAAG TGAAAGAGTT CACTCCAAAT CAGTAGAGAG TAATATTGAA GACAAAATAT 1500

TTGGGAAAAC CTATCGGAAG AAGGCAAGCC TCCCCAACTT AAGCCATGTA ACTGAAAATC 1560

TAATTATAGG AGCATTTGTT ACTGAGCCAC AGATAATACA AGAGCGTCCC CTCACAAATA 1620

AATTAAAGCG TAAAAGGAGA CCTACATCAG GCCTTCATCC TGAGGATTTT ATCAAGAAAG 1680

CAGATTTGGC AGTTCAAAAG ACTCCTGAAA TGATAAATCA GGGAACTAAC CAAACGGAGC 1740

AGAATGGTCA AGTGATGAAT ATTACTAATA GTGGTCATGA GAATAAAACA AAAGGTGATT 1800

CTATTCAGAA TGAGAAAAAT CCTAACCCAA TAGAATCACT CGAAAAAGAA TCTGCTTTCA 1860

AAACGAAAGC TGAACCTATA AGCAGCAGTA TAAGCAATAT GGAACTCGAA TTAAATATCC 1920

ACAATTCAAA AGCACCTAAA AAGAATAGGC TGAGGAGGAA GTCTTCTACC AGGCATATTC 1980

ATGCGCTTGA ACTAGTAGTC AGTAGAAATC TAAGCCCACC TAATTGTACT GAATTGCAAA 2040

TTGATAGTTG TTCTAGCAGT GAAGAGATAA AGAAAAAAAA GTACAACCAA ATGCCAGTCA 2100

GGCACAGCAG AAACCTACAA CTCATGGAAG GTAAAGAACC TGCAACTGGA GCCAAGAAGA 2160

GTAACAAGCC AAATGAACAG ACAAGTAAAA GACATGACAG CGATACTTTC CCAGAGCTGA 2220

AGTTAACAAA TGCACCTGGT TCTTTTACTA AGTGTTCAAA TACCAGTGAA CTTAAAGAAT 2280

TTGTCAATCC TAGCCTTCCA AGAGAAGAAA AAGAAGAGAA ACTAGAAACA GTTAAAGTGT 2340

CTAATAATGC TGAAGACCCC AAAGATCTCA TGTTAAGTGG AGAAAGGGTT TTGCAAACTG 2400

AAAGATCTGT AGAGAGTAGC AGTATTTCAT TGGTACCTGG TACTGATTAT GGCACTCAGG 2460

AAAGTATCTC GTTACTGGAA GTTAGCACTC TAGGGAAGGC AAAAACAGAA CCAAATAAAT 2520

GTGTGAGTCA GTGTGCAGCA TTTGAAAACC CCAAGGGACT AATTCATGGT TGTTCCAAAG 2580

ATAATAGAAA TGACACAGAA GGCTTTAAGT ATCCATTGGG ACATGAAGTT AACCACAGTC 2640

GGGAAACAAG CATAGAAATG GAAGAAAGTG AACTTGATGC TCAGTATTTG CAGAATACAT 2700

TCAAGGTTTC AAAGCGCCAG TCATTTGCTC CGTTTTCAAA TCCAGGAAAT GCAGAAGAGG 2760

AATGTGCAAC ATTCTCTGCC CACTCTGGGT CCTTAAAGAA ACAAAGTCCA AAAGTCACTT 2820

TTGAATGTGA ACAAAAGGAA GAAAATCAAG GAAAGAATGA GTCTAATATC AAGCCTGTAC 2880

AGACAGTTAA TATCACTGCA GGCTTTCCTG TGGTTGGTCA GAAAGATAAG CCAGTTGATA 2940

ATGCCAAATG TAGTATCAAA GGAGGCTCTA GGTTTTGTCT ATCATCTCAG TTCAGAGGCA 3000

ACGAAACTGG ACTCATTACT CCAAATAAAC ATGGACTTTT ACAAAACCCA TATCGTATAC 3060

CACCACTTTT TCCCATCAAG TCATTTGTTA AAACTAAATG TAAGAAAAAT CTGCTAGAGG 3120

AAAACTTTGA GGAACATTCA ATGTCACCTG AAAGAGAAAT GGGAAATGAG AACATTCCAA 3180

GTACAGTGAG CACAATTAGC CGTAATAACA TTAGAGAAAA TGTTTTTAAA GAAGCCAGCT 3240

CAAGCAATAT TAATGAAGTA GGTTCCAGTA CTAATGAAGT GGGCTCCAGT ATTAATGAAA 3300

TAGGTTCCAG TGATGAAAAC ATTCAAGCAG AACTAGGTAG AAACAGAGGG CCAAAATTGA 3360

ATGCTATGCT TAGATTAGGG GTTTTGCAAC CTGAGGTCTA TAAACAAAGT CTTCCTGGAA 3420

GTAATTGTAA GCATCCTGAA ATAAAAAAGC AAGAATATGA AGAAGTAGTT CAGACTGTTA 3480

ATACAGATTT CTCTCCATAT CTGATTTCAG ATAACTTAGA ACAGCCTATG GGAAGTAGTC 3540

ATGCATCTCA GGTTTGTTCT GAGACACCTG ATGACCTGTT AGATGATGGT GAAATAAAGG 3600

AAGATACTAG TTTTGCTGAA AATGACATTA AGGAAAGTTC TGCTGTTTTT AGCAAAAGCG 3660

TCCAGAAAGG AGAGCTTAGC AGGAGTCCTA GCCCTTTCAC CCATACACAT TTGGCTCAGG 3720

GTTACCGAAG AGGGGCCAAG AAATTAGAGT CCTCAGAAGA GAACTTATCT AGTGAGGATG 3780

AAGAGCTTCC CTGCTTCCAA CACTTGTTAT TTGGTAAAGT AAACAATATA CCTTCTCAGT 3840

CTACTAGGCA TAGCACCGTT GCTACCGAGT GTCTGTCTAA GAACACAGAG GAGAATTTAT 3900

TATCATTGAA GAATAGCTTA AATGACTGCA GTAACCAGGT AATATTGGCA AAGGCATCTC 3960

AGGAACATCA CCTTAGTGAG GAAACAAAAT GTTCTGCTAG CTTGTTTTCT TCACAGTGCA 4020

GTGAATTGGA AGACTTGACT GCAAATACAA ACACCCAGGA TCCTTTCTTG ATTGGTTCTT 4080

CCAAACAAAT GAGGCATCAG TCTGAAAGCC AGGGAGTTGG TCTGAGTGAC AAGGAATTGG 4140

TTTCAGATGA TGAAGAAAGA GGAACGGGCT TGGAAGAAAA TAATCAAGAA GAGCAAAGCA 4200

TGGATTCAAA CTTAGGTGAA GCAGCATCTG GGTGTGAGAG TGAAACAAGC GTCTCTGAAG 4260

ACTGCTCAGG GCTATCCTCT CAGAGTGACA TTTTAACCAC TCAGCAGAGG GATACCATGC 4320

AACATAACCT GATAAAGCTC CAGCAGGAAA TGGCTGAACT AGAAGCTGTG TTAGAACAGC 4380

ATGGGAGCCA GCCTTCTAAC AGCTACCCTT CCATCATAAG TGACTCTTCT GCCCTTGAGG 4440

ACCTGCGAAA TCCAGAACAA AGCACATCAG AAAAAGCAGT ATTAACTTCA CAGAAAAGTA 4500

GTGAATACCC TATAAGCCAG AATCCAGAAG GCCTTTCTGC TGACAAGTTT GAGGTGTCTG 4560

CAGATAGTTC TACCAGTAAA AATAAAGAAC CAGGAGTGGA AAGGTCATCC CCTTCTAAAT 4620

GCCCATCATT AGATGATAGG TGGTACATGC ACAGTTGCTC TGGGAGTCTT CAGAATAGAA 4680

ACTACCCATC TCAAGAGGAG CTCATTAAGG TTGTTGATGT GGAGGAGCAA CAGCTGGAAG 4740

AGTCTGGGCC ACACGATTTG ACGGAAACAT CTTACTTGCC AAGGCAAGAT CTAGAGGGAA 4800

CCCCTTACCT GGAATCTGGA ATCAGCCTCT TCTCTGATGA CCCTGAATCT GATCCTTCTG 4860

AAGACAGAGC CCCAGAGTCA GCTCGTGTTG GCAACATACC ATCTTCAACC TCTGCATTGA 4920

AAGTTCCCCA ATTGAAAGTT GCAGAATCTG CCCAGAGTCC AGCTGCTGCT CATACTACTG 4980

ATACTGCTGG GTATAATGCA ATGGAAGAAA GTGTGAGCAG GGAGAAGCCA GAATTGACAG 5040

CTTCAACAGA AAGGGTCAAC AAAAGAATGT CCATGGTGGT GTCTGGCCTG ACCCCAGAAG 5100

AATTTATGCT CGTGTACAAG TTTGCCAGAA AACACCACAT CACTTTAACT AATCTAATTA 5160

CTGAAGAGAC TACTCATGTT GTTATGAAAA CAGATGCTGA GTTTGTGTGT GAACGGACAC 5220

TGAAATATTT TCTAGGAATT GCGGGAGGAA AATGGGTAGT TAGCTATTTC TGGGTGACCC 5280

AGTCTATTAA AGAAAGAAAA ATGCTGAATG AGCATGATTT TGAAGTCAGA GGAGATGTGG 5340

TCAATGGAAG AAACCACCAA GGTCCAAAGC GAGCAAGAGA ATCCCAGGAC AGAAAGATCT 5400

TCAGGGGGCT AGAAATCTGT TGCTATGGGC CCTTCACCAA CATGCCCACA GATCAACTGG 5460

AATGGATGGT ACAGCTGTGT GGTGCTTCTG TGGTGAAGGA GCTTTCATCA TTCACCCTTG 5520

GCACAGGTGT CCACCCAATT GTGGTTGTGC AGCCAGATGC CTGGACAGAG GACAATGGCT 5580

TCCATGCAAT TGGGCAGATG TGTGAGGCAC CTGTGGTGAC CCGAGAGTGG GTGTTGGACA 5640

GTGTAGCACT CTACCAGTGC CAGGAGCTGG ACACCTACCT GATACCCCAG ATCCCCCACA 5700

GCCACTACTG A 5711

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Consensus sequence"

(iii) HYPOTHETICAL: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GGTCANNNTG GTCNNNNNNN NNTGACC 27

WHAT IS CLAIMED IS:

1. A DNA molecule comprising a sequence shown as SEQ ID NO:2.

2. A DNA molecule comprising a sequence complementary to SEQ ID NO:2.

3. An RNA molecule complementary to the DNA of claim 1.

4. An RNA molecule complementary to the DNA of claim 2.

5. A nucleic acid consisting essentially of the sequence shown by SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5 or SEQ ID NO:6.

6. A nucleic acid consisting essentially of the sequence shown by SEQ ID NO:7 or SEQ ID NO:8.

7. A nucleic acid consisting essentially of the sequence shown by SEQ ID NO:10.

8. A method for specifically amplifying a portion of the *BRCA1* gene or cDNA while not amplifying the *LBRCA1* gene or cDNA, said method comprising performing a polymerase chain reaction using primers 120.2 and 120.3 and using cycling conditions consisting essentially of 45 seconds at 94°C, 60 seconds at 57°C, and 90 seconds at 72°C.

9. A method for specifically amplifying a portion of the *BRCA1* gene or cDNA while not amplifying the *LBRCA1* gene or cDNA, said method comprising performing a polymerase chain reaction using primers 214.3 and 42.2 and using a first set of cycling conditions followed by a second set of cycling conditions wherein said first set of cycling conditions consists essentially of cycles of 30 seconds at 94°C, 60 seconds at 60°C, and 165 seconds at 72°C and said second set of cycling conditions consists essentially of cycles of 30 seconds at 94°C, 60 seconds at 55°C and 170 seconds at 72°C.

10. A method for specifically amplifying a portion of the *BRCA1* gene or cDNA while not amplifying the *LBRCA1* gene or cDNA, said method comprising performing a polymerase chain reaction using primers 42.2 and 120.2 and using cycling conditions consisting essentially of 30 seconds at 94°C, 60 seconds at 58°C and 90 seconds at 72°C.

11. A method for specifically amplifying a portion of the *1A1.3B* gene or cDNA while not amplifying the *L1A1.3B* gene or cDNA, said method comprising performing a polymerase chain reaction using primers 225.1 and 225.4 and using cycling conditions consisting essentially of 45 seconds at 94°C, 60 seconds at 61°C and 90 seconds at 72°C.

12. A method for analyzing somatic tissue for deletion of at least a portion of BRCA1 from one chromosome in a person who is heterozygous for a polymorphism, said method consisting of the following steps:

(a) determining whether the person is heterozygous in germline tissue for a specific polymorphism within BRCA1 or its promoter region;

(b) determining whether the person is heterozygous in said somatic tissue for said specific polymorphism; and

(c) comparing the zygosity of the polymorphism in said germline tissue and said somatic tissue wherein:

1) if the germline tissue is heterozygous and the somatic tissue is heterozygous for the polymorphism then there has been no deletion of the polymorphic gene region;

2) if the germline tissue is heterozygous and the somatic tissue is not heterozygous then there has been a deletion of the polymorphic gene region; and

3) if the germline tissue is homozygous the assay is uninformative unless said somatic tissue is null for the polymorphism thereby indicating a loss of all copies of the gene region within the somatic tissue.

13. The method of claim 12 wherein said polymorphism is the C/T polymorphism at base 612 of SEQ ID NO:1

14. The method of claim 12 wherein said polymorphism is the AAC/AACAAC polymorphism at bases 980-982 of SEQ ID NO:1.

15. The method of claim 12 wherein said polymorphism is the A/G polymorphism at base 1723 of SEQ ID NO:2.

16. The method of claim 12 wherein said polymorphism is the A/G polymorphism at base 2182 of SEQ ID NO:2.

17. A method for determining the copy number of BRCA1 genes within a human genome by using a quantitative polymerase chain reaction.

18. The method of claim 17 wherein PCR primers corresponding to a fragment of SEQ ID NO:2 or its complement are used.

19. A method for determining the copy number and large-scale genomic structure of a human genomic region containing a BRCA1 promoter using pulsed-field gel electrophoresis.

20. A method for specifically amplifying a target nucleic acid that comprises at least 25 consecutive nucleotides of SEQ ID NO:1 or its complement while not amplifying a second nucleic acid that comprises at least 25 consecutive nucleotides of SEQ ID NO:2 or its complement, wherein said method comprises performing a polymerase chain reaction using primers with 3' termini wherein when said primers hybridize to said target nucleic acid said 3' termini will be complementary to a strand of said target nucleic acid to which said primer hybridizes and wherein if said primers bind to said second nucleic acid said 3' termini will not be complementary to a strand of said second nucleic acid to which said primer binds.

21. The method of claim 20 wherein said 3' termini are defined as a single nucleotide which is at the ultimate 3' position of each primer.

22. The method of claim 20 wherein said 3' termini are defined as two nucleotides which are the final two nucleotides at the 3' end of each primer.

23. A nucleic acid comprising at least 10 consecutive nucleotides of SEQ ID NO:2 or its complement.

24. A method of performing a polymerase chain reaction wherein said method uses primers that have a nucleotide sequence identical to a portion of SEQ ID NO:2 or its complement.

25. Nucleic acid oligonucleotides useful as primers for a polymerase chain reaction wherein said oligonucleotides consist of a nucleic acid sequence that is identical to a portion of SEQ ID NO:2 or its complement.

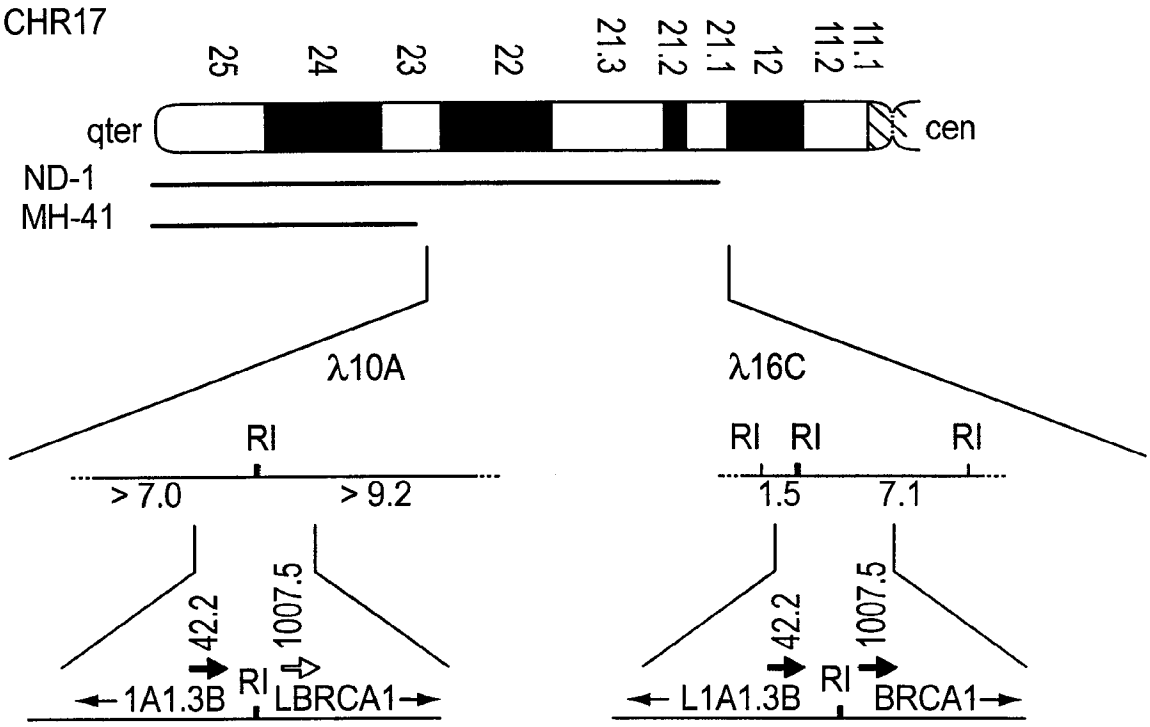


FIG. 1A

		RNU2	1A1.3B exon 19	1A1.3B exon 12	1A1.3B promoter	BRCA1 promoter	BRCA1 exon11
LIBRARY	CLONE				225.1+ 225.4	120.2+ 42.2	
L17 LANL	16C	−	−	−	−	+	−
P1	746B4	−	−	−	−	+	+
YAC CEPH	167B7	−	−	−	+	+	−
L17 LANL	10A	−	−	−	+	−	−
YAC CEPH	173B7	+	+	+	+	−	−
YAC St. L	A167E6	+	+	+	−	−	−

FIG. 1B

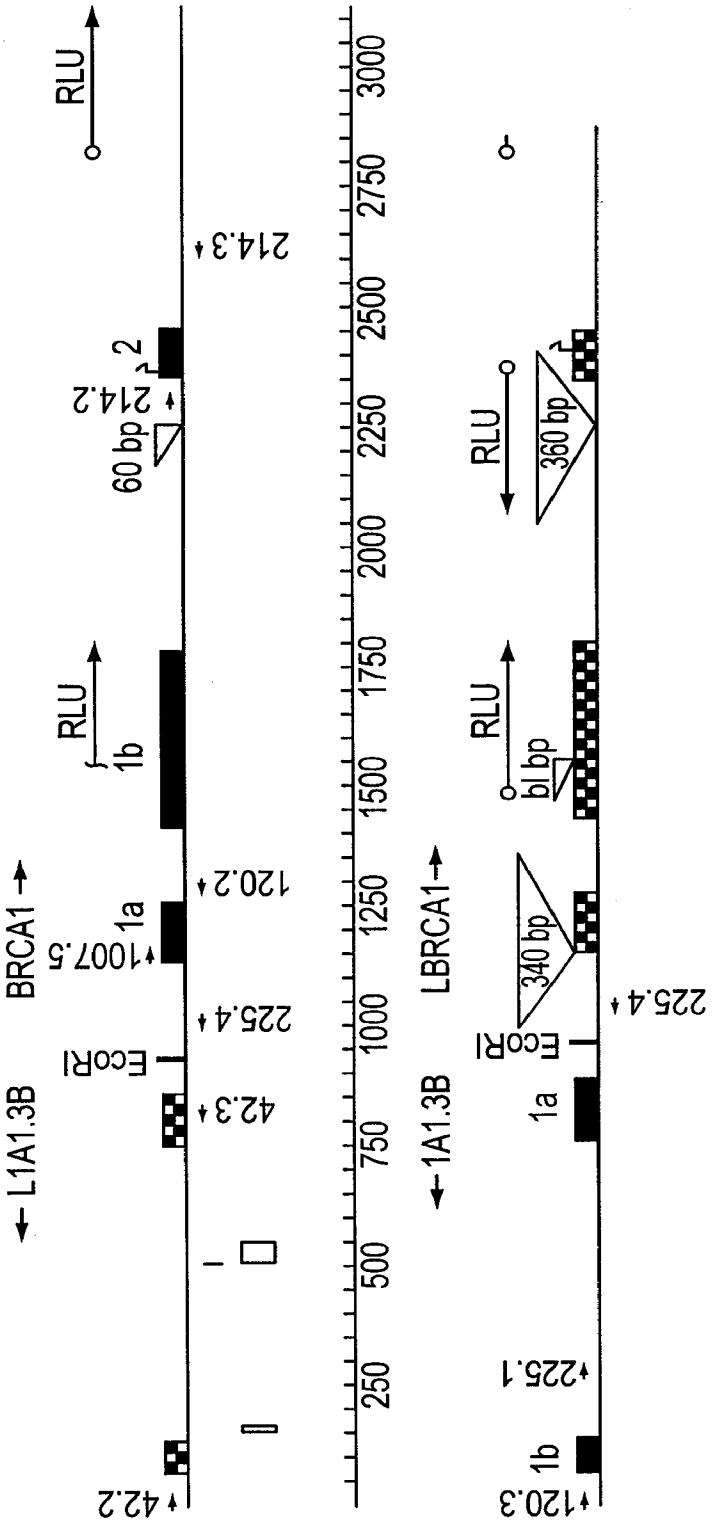

RLU
214.3
214.2
2
60 bp
RLU
1b
120.2
1a
1007.5
225.4
EcoRI
42.3
L1A1.3B
BRCA1
42.2
250
500
750
1000
1250
1500
1750
2000
2250
2500
2750
3000
RLU
360 bp
RLU
81 bp
340 bp
225.4
EcoRI
1a
1A1.3B
LBRCA1
225.1
1b
120.3

FIG. 2

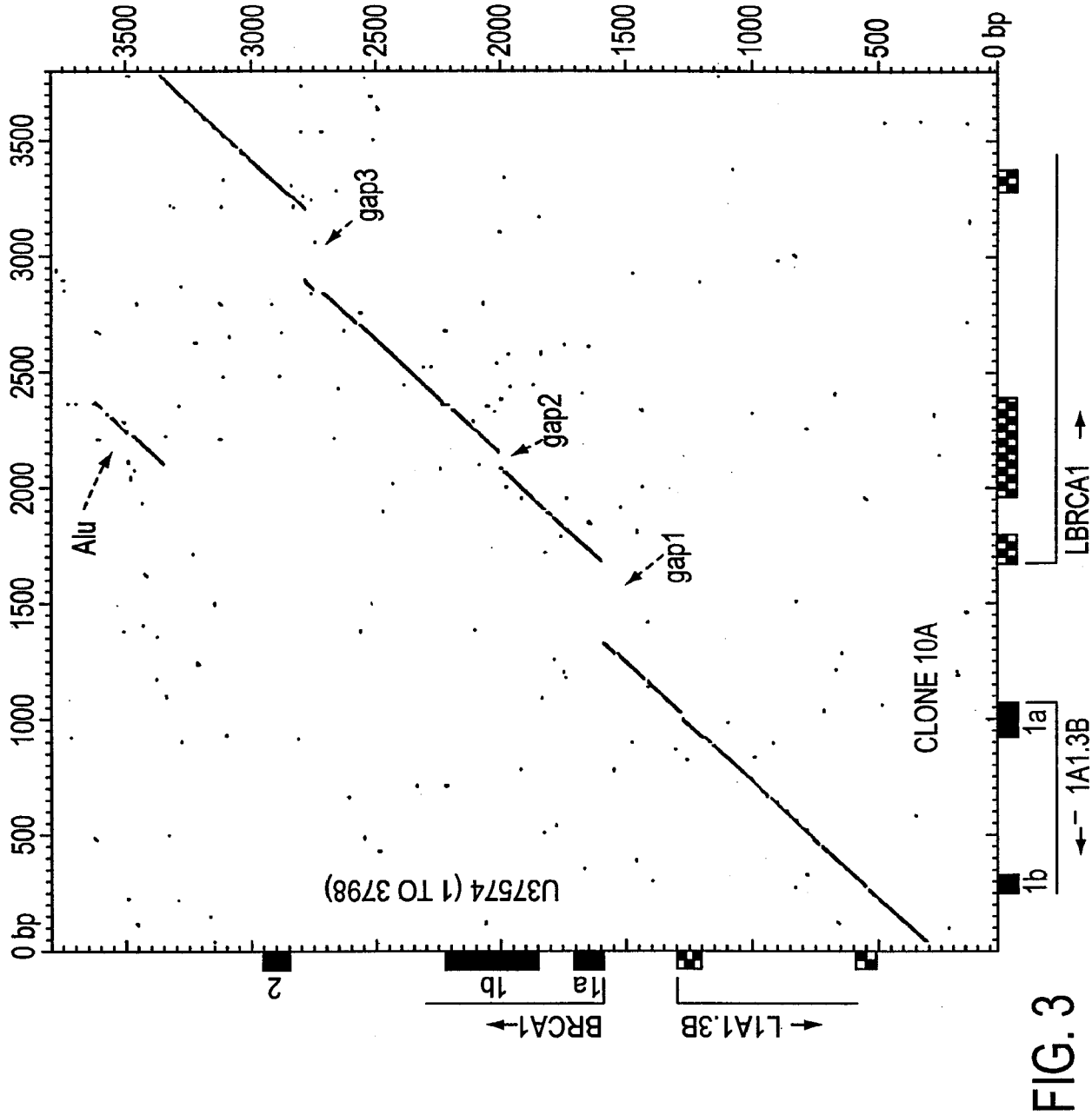

0 bp
500
1000
1500
2000
2500
3000
3500
gap3
gap2
gap1
Alu
U37574 (1 TO 3798)
CLONE 10A
LBRCA1
1a
1b
1A1.3B
L1A1.3B
BRCA1
1a
1b
2

FIG. 3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/21358

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C12Q 1/68; C12P 19/34; C07H 21/04
US CL :435/6, 91.2; 536/23.5, 24.31, 24.33

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 91.2; 536/23.5, 24.31, 24.33

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	BROWN et al. The 5' end of the BRCA1 gene lies within a duplicated region of human chromosome 17q21. Oncogene. June 1996, Vol. 12, No. 12, pages 2507-2513, especially figure 4 and page 2511.	1-25
Y,P	BARKER et al. The BRCA1 and 1A1.3B promoters are parallel elements of a genomic duplication at 17q21. Genomics. December 1996, Vol. 38, pages 215-222, especially pages 215-216.	1-25
Y	CAMPBELL et al. A novel gene encoding a B-box protein within the BRCA1 region at 17q21.1. Human Molecular Genetics. April 1994, Vol. 3, No. 4, pages 589-594, especially figure 1 and page 593.	1-4, 6, 11, 20-25

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 10 MARCH 1998	Date of mailing of the international search report 03 APR 1998
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer CARLA MYERS Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/21358

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Database GenBank on STN, Accession no. U37574, Xu et al, 1996.	5, 7

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/21358

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

EMBL; GenBank; N-GENE-SEQ; APS; DIALOG: Medline, CA, Derwent Patents, Biosis
search terms: SEQ ID NO: 1-8, 10; BRCA1, 1A1.3B, LBRCA1, L1A1.3B, mutation, copy number, polymorphism